

SVEUČILIŠTE U ZAGREBU
FILOZOFSKI FAKULTET
ODSJEK ZA INFORMACIJSKE I KOMUNIKACIJSKE
ZNANOSTI
Akademska godina 2015./2016.

Astrid Plišo
UMJETNA OPĆA INTELIGENCIJA
Diplomski rad

Mentor: dr. sc. Vjera Lopina

Zagreb, 2016.

SADRŽAJ

1. UVOD	1
2. KRATKA POVIJEST	2
3. DEFINICIJA UMJETNE OPĆE INTELIGENCIJE	5
3.1. Pristupi definiranju inteligencije	6
3.1.1. Psihološki pristup	6
3.1.2. Pragmatični pristup	8
3.1.3. Pristup teorije kontrole	9
3.1.4. Efikasna inteligencija	10
3.1.5. Utjelovljeni pristup	11
3.2. Tipologija UOI sustava	12
3.2.1. Struktura	12
3.2.2. Ponašanje	13
3.2.3. Sposobnost	13
3.2.4. Funkcija	14
3.2.5. Načelo	14
4. IZRADA UOI SUSTAVA	16
4.1. Simbolički pristup	16
4.1.1. Prikaz znanja	17
4.1.2. Učenje	20
4.1.2.1. CYC	22
4.1.2.2. SOAR	22
4.1.2.3. EPIC	23
4.2. Podsimbolički pristup	24
4.2.1. Prikaz znanja	26
4.2.2. Učenje	27
4.2.2.1. IBCA	31
4.2.2.2. CORCOTRONICS	32
4.2.2.3. DeSTIN	32
4.3. Hibridni pristup	33
4.3.1. ACT-R	36
4.3.2. CLARION	37
4.3.3. LIDA	38
5. ARGUMENTI PROTIV UOI SUSTAVA	39
5.1. UI kao aksiomatski sustav	39
5.2. UI i Turingov stroj	40
5.3. UI i pitanje semantike	41
5.4. Ostale kritike	42
6. Zaključak	44
7. Literatura	45

1. UVOD

Područje umjetne inteligencije jedna je od najdinamičnijih znanstvenih disciplina 21. stoljeća. Njezine brojne ideje i praktični izumi koji se danas koriste u područjima financija, medicine, telekomunikacije, avijacije, računalne znanosti itd., radikalno su izmijenili svakodnevni život, iako većina ljudi toga nije ni svjesna. No unatoč fascinatnim dostignućima, ovakva primjena umjetne inteligencije predstavlja veliki odmak od njezinog prvotnog cilja. Naime, originalna vizija ovoga područja oduvijek je bila izgradnja računalnih sustava i arhitektura koji će posjedovati inteligenciju ravnu ljudskoj te ju u konačnici i premašiti. Još tijekom 60-ih godina prošlog stoljeća najave o istinski inteligentnim programima koji će postati dio svakodnevice smatrane su kao sasvim realna i ostvariva mogućnost. No suočeni s enormnom kompleksnošću ljudske inteligencije, kao i gubitkom financijskih izvora za svoje projekte koji u konačnici nisu pokazali nikakvu praktičnu svrhu, većina UI znanstvenika i istraživača polako je odustajala od ovog cilja. Nova dominantna paradigma unutar umjetne inteligencije postala je izrada usko specijaliziranih programa, a kriteriji uspješnosti mogućnost njihove praktične primjene i isplativost. Ideja o pravoj umjetnoj inteligenciji uvelike je zaboravljena iako su određeni znanstvenici inzistirali kako će ovaj novi pristup prije ili kasnije rezultirati pojavom istinskih UI agenata.

Razočarani ovakvim reduciranim viđenjem područja umjetne inteligencije, određeni broj istraživača posljednjih godina polako se vraća prvotnom idealu. U posljednjem desetljeću nastao je veliki broj konkretnih programa, radionica, konferencija i specijaliziranih časopisa usmjerenih prema teoretskoj i praktičnoj razradi istinski inteligentnih sustava. Kako bi naglasili fundamentalnu karakteristiku ovakvog pristupa odabran je naziv umjetna opća inteligencija (*Artificial general intelligence - AGI*). Osnovni cilj ovoga rada je prikazati glavne teoretske odrednice koje razlikuju ovaj pristup od ostatka tradicionalne umjetne inteligencije, a naglasak je na problemu definiranja pojma ljudske ili opće inteligencije. Također, u narednim poglavljima biti će prikazani osnovni pristupi u izradi takvih sustava s fokusom na njihovim prednostima i manama sa stajališta zahtjeva umjetne opće inteligencije. Bitno je napomenuti kako je suština ovoga rada usmjerena na razne konceptualne i praktične prepreke u ostvarenju ove ideje. Konkretno, to znači da se ne spominju razne filozofske i metafizičke kategorije poput posjedovanja svijesti, empatije ili etičke implikacije izgradnje inteligentnih sustava. Iako su takve teme od iznimne važnosti, one ipak spadaju pod sasvim drugi aspekt umjetne inteligencije.

2. KRATKA POVIJEST

Iako povijest nije u središtu zanimanja ovoga rada nužno je pružiti kratki pregled razvoja ideje umjetne inteligencije radi boljeg shvaćanja cjelokupnog područja. Naravno, sama ideja nije novijeg porijekla nego vuče svoje korijene još iz antičkog razdoblja, iz raznih mitova i legendi poput Hefestovih robota ili Talosa s Krete, dok je samu teoretsku podlogu u vidu formalnog rasuđivanja moguće pronaći još u razmišljanjima antičkih filozofa kao što su Aristotel i Euklid. Kao početak moderne znanosti umjetne inteligencije uglavnom se navodi 1950. godina kada je Alan Turing izdao svoj članak *Computing Machinery and Intelligence* te formulirao sada već famozni Turingov test. Nedugo nakon toga dolazi do razvoja brojnih ideja koje su imale presudan utjecaj na razvoj cjelokupnog područja. Neke od njih su teorija kibernetike Norberta Wienera, informacijska teorija Claudea Shannona te umjetni neuroni Waltera Pittsa i Warrena McCullocha. Kao primjer prvog inteligentnog programa navodi se *Logic Theorist* Allena Newella i Herberta Simona izrađenog 1955. godine za potrebe dokazivanja matematičkih teorema. Ovo početno razdoblje doživljava svoj vrhunac 1956. godine s konferencijom u Dartmouthu gdje je službeno zaključeno kako se svaki aspekt inteligencije može računalno simulirati. Na inicijativu Johna McCarthyja ovo područje dobiva naziv umjetne inteligencije.¹

Nakon konferencije uslijedila je prava eksplozija teoretskih istraživanja i praktičnih programa koja je potrajala sve do 1974. godine. U ovome periodu koncept heuristike se uspostavlja kao jedan od osnovnih principa funkcioniranja inteligentnih programa. Neki od važnijih događaja ovog razdoblja su nastanak specijaliziranog programskog jezika LISP koji se koristi i danas te izrada ELIZA-e, prvog svjetskog chatterbota koji je bio sposoban simulirati ljudske razgovore.² Nažalost, ovo razdoblje velikog optimizma uskoro je doživjelo potpuni krah. Postalo je jasno kako će brojna očekivanja i obećanja vodećih znanstvenika ostati neispunjena. Kao osnovni problem su identificirana tadašnja ograničenja u računalnoj memoriji i procesorskoj snazi. Ovakva ograničenja imala su najviše utjecaja u odnosu na takozvani Moravecov paradoks. Naime, Hans Moravec je prvi formulirao jednu od osnovnih zakonitosti cjelokupnog UI područja koja kaže kako niskorazinske senzo-motoričke vještine

¹ History of artificial intelligence. Wikipedia.

URL:[http://en.wikipedia.org/wiki/History_of_artificial_intelligence#Dartmouth_Conference_1956:_the_birt
h_of_AI](http://en.wikipedia.org/wiki/History_of_artificial_intelligence#Dartmouth_Conference_1956:_the_birth_of_AI) (20.8.2016.)

² Artificial intelligence, Evolutionary computing. Encyclopaedia Britannica.

URL:[http://www.britannica.com/EBchecked/topic/37146/artificial-intelligence-AI/219093/Evolutionary-
computing](http://www.britannica.com/EBchecked/topic/37146/artificial-intelligence-AI/219093/Evolutionary-computing) (20.8.2016.)

zahtijevaju puno veće količine računalnih resursa od visokorazinskih procesa.³ Laički rečeno, shvatilo se da je relativno jednostavno osposobiti računala za igranje šaha dok su se sposobnosti percepcije ili mobilnosti pokazale nerješivima. Ovo razdoblje nesigurnosti i konflikta rezultiralo je velikom podjelom znanstvenika i istraživača na dvije dominantne struje. Prva je nazvana „*neats*“ te je zastupala isključivu uporabu formalnih metoda poput logike i statistike u programiranju umjetne inteligencije, a kao njezin predvodnik istaknuo se John McCarthy. Druga struja, nazvana „*scruffies*“, smatrala je kako se ljudska inteligencija ne može adekvatno simulirati korištenjem samo formalnih metoda te su inzistirali na određenoj dozi kreativnosti i intuitivnom prilagođavanju programa. Veliki zagovornik ovog pristupa bio je Marvin Minsky.⁴ Kao zgodan detalj vrijedi napomenuti kako je druga skupina zaslužna za nastanak hakerske subkulture.

Oživljavanje cjelokupnog područja umjetne inteligencije nastupilo je oko 1980. godine. U ovome periodu dolazi do potpune promjene dominantne paradigme UI. Naime, osnovni cilj ovoga područja od početka je bio dizajn i izrada strojeva koji će posjedovati razinu inteligencije usporedivu s ljudima. No sve se promijenilo sa spektakularnim uspjehom ekspertnih sustava počevši s XCON sustavom razvijenim na Carnegie Mellon sveučilištu. Jednostavnom kombinacijom znanja raznih stručnjaka i logičkih pravila shvatilo se kako UI može biti itekako efikasna u usko definiranim područjima.⁵ Znanstvenici su se napokon mogli pohvaliti s praktičnom primjenom svojih izuma, a efikasnost je postala nova misao vodilja područja umjetne inteligencije. Ovaj novi koncept izrade specijaliziranih programa namijenjenih za uporabu u striktno definiranim domenama postao je poznat kao uska UI (*narrow AI*) te je na snazi i danas. Unatoč uspjehu ekspertnih sustava, njihova primjena uskoro se pokazala previše ograničavajućom, a početni entuzijizam ubrzo je splasnulo. Inteligentni programi opet su se pokazali kao nerealna fantazija znanstvenika, a cjelokupno područje UI još jednom se našlo u određenoj krizi identiteta. Izlazak iz te krize pružilo je područje robotike. Točnije, istraživanja Davida Marra kojima je odbacio teoretske postavke McCarthyja i njegovo inzistiranje na ekskluzivnom korištenju formalnih metoda. Naime, Marr je tvrdio kako su osnovne senzo-motoričke vještine neophodne za sve ostale visoko razinske vještine kao što je zdravorazumsko razmišljanje. Prema Marru, ljudi prilikom izvršavanja ovih osnovnih vještina ne koriste nikakvu manipulaciju simbola što pak znači da su formalne

³ Moravec's paradox. Wikipedia. URL: http://en.wikipedia.org/wiki/Moravec's_paradox (20.8.2016.)

⁴ History of artificial intelligence. Wikipedia.

URL: [http://en.wikipedia.org/wiki/History_of_artificial_intelligence#The_neats: logic.2C_Prolog_and_expert_systems](http://en.wikipedia.org/wiki/History_of_artificial_intelligence#The_neats:_logic.2C_Prolog_and_expert_systems) (20.8.2016.)

⁵ Expert system. Wikipedia. URL: http://en.wikipedia.org/wiki/Expert_system (20.8.2016.)

metode programiranja UI u ovim situacijama beskorisne. Brojni istraživači, poput Rodneya Brooksa, uskoro su se priklonili ovakvom tumačenju tvrdeći kako je nekakav oblik tijela potreban za bilo kakvu vrstu spoznaje. Ova dihotomija između simboličkog i podsimboličkog programiranja umjetne inteligencije u određenoj mjeri zadržala se do danas.

Za današnju modernu eru UI uglavnom se tvrdi kako je započela 1993. godine. Za njezine brojne uspjehe pretežito je zaslužan tehnološki napredak u vidu konstantnog povećanja brzine i kapaciteta računala prema poznatom Mooreovom zakonu. Iako je umjetna inteligencija doživjela veliki komercijalni uspjeh, cjelokupno područje prošlo je kroz znatnu fragmentaciju na brojna različita područja i smjerove istraživanja, često međusobno nepovezana. No jedna od pozitivnih nuspojava ovoga je konvergencija ideja iz različitih područja te uspostava novih paradigmi. Jedan takav primjer bili bi inteligentni agenti (*intelligent agents*), nastali spajanjem koncepata iz ekonomije i teorije odluka s umjetnom inteligencijom. Ove agente karakterizira sposobnost percepcije svoje okoline te donošenje optimalnih odluka na temelju nje kako bi maksimizirali šanse za postizanje svojih ciljeva.⁶ Također, jedna od specifičnosti ovoga razdoblja je uspostavljanje matematike i logike kao dominantnih metoda programiranja UI. Jedna od prednosti ovakvog matematičkog pristupa je bolja suradnja između različitih područja, ali i mogućnost konkretnog mjerenja dobivenih rezultata. Ovime je područje UI postalo rigoroznija znanstvena disciplina, a mnogi su ovo protumačili kao konačnu pobjedu filozofije *neatsa*. Uostalom, ovakav pristup savršeno je u skladu s prioritetima uske UI, a to je izrada efikasnih sustava koje je moguće koristiti i evaluirati u svakodnevnim situacijama. Unatoč ovome, prvotni ideal izrade istinski inteligentnih strojeva i dalje je živ iako je pod pritiskom komercijalne isplativosti UI donedavno bio posve zanemaren. Naravno svi oni problemi koji su onemogućavali ostvarenje tog ideala i dalje su prisutni, a među njima se svakako izdvaja nepostojanje jedinstvene definicije same inteligencije prema kojoj bi UI sustav trebao biti izrađen. U narednim poglavljima biti će prikazane neke od mogućih definicija inteligencije kao i glavni pristupi dizajnu i izradi umjetno inteligentnih programa i arhitektura.

⁶ Intelligent agent. Wikipedia. URL: http://en.wikipedia.org/wiki/Intelligent_agent (20.8.2016.)

3. DEFINICIJA UMJETNE OPĆE INTELIGENCIJE

Za početak je nužno pružiti nekakvu deskripciju umjetne opće inteligencije. Nažalost, kao što je to često slučaj u ovome području, ne postoji jedna opće prihvaćena definicija. Njezine karakteristike možda se najbolje očituju u usporedbi s uskom UI. Naime, velika većina današnjih sustava dizajnirani su za imitaciju inteligentnog ponašanja u veoma specifičnim kontekstima. Prilikom najmanje promjene tog konteksta ovakvi sustavi zahtijevaju neku vrstu reprogramiranja ili rekonfiguracije kako bi zadržali svoju razinu inteligencije. S druge strane UOI sustavi trebali bi posjedovati široku sposobnost samostalne prilagodbe na promjene u svojim ciljevima i okolnostima, odnosno moraju biti sposobni generalizirati znanje iz jednog cilja ili konteksta u drugi. Jedna bitna distinkcija je da sustav ne mora posjedovati beskonačnu općenitost, prilagodljivost ili fleksibilnost kako bi se smatrao općom inteligencijom. Iako je nemoguće pronaći jedinstvenu definiciju, unutar područja postoji konsenzus o određenim kvalitativnim karakteristikama ovakvih sustava:

- Opća inteligencija uključuje sposobnost postizanja različitih ciljeva te obavljanja raznih zadataka u različitim kontekstima i okruženjima
- Opće inteligentni sustav mora se snalaziti u problemima i situacijama sasvim različitim od onih koje su predvidjeli njegovi kreatori
- Ovakvi sustavi trebali bi biti dobri u generaliziranju stečenog znanja kako bi mogli prenijeti to znanje iz jednog problema ili konteksta u drugi
- Proizvoljna opća inteligencija nije moguća s obzirom na realno ograničenje resursa
- U stvarnom svijetu sustavi mogu prikazivati različite stupnjeve općenitosti, ali će neizbježno biti učinkovitiji u učenju određenih stvari. To znači da su u stvarnom svijetu opće inteligencije pristrane određenim vrstama ciljeva i okruženja
- Ljudi pokazuju višu razinu opće inteligencije od postojećih UI programa
- Čini se malo vjerojatnim da ljudi posjeduju maksimalnu razinu opće inteligencije, čak i u odnosu na ciljeve i okruženje za koje su evolucijski prilagođeni⁷

Sam pojam umjetne opće inteligencije relativno je noviji. Ovo područje je dugi niz godina bilo poznato kao jaka UI, ali je taj izraz s vremenom postao nepoželjan zbog toga što ga je John Searle često koristio u svojim argumentima protiv umjetne inteligencije kao što je

⁷ Goertzel, B. Artificial General Intelligence: Concept, State of the Art, and Future Prospects, u: *Journal of Artificial General Intelligence*, vol. 5, br. 1, 15. ožujka 2014. Str. 1-48.
URL: <https://www.degruyter.com/downloadpdf/j/jagi.2014.5.issue-1/jagi-2014-0001/jagi-2014-0001.xml> (20.8.2016.)

poznata kineska soba. Ponekad se također koristi i izraz ljudska UI (*human-level AI*), no ovaj pojam se također često izbjegava zbog nepostojanja univerzalne definicije ljudske inteligencije. U konačnici, sama svrha pojma i koncepta umjetne opće inteligencije je usmjeravanje pažnje na opći opseg i sposobnost generalizacije određenih inteligentnih sustava, kao što su ljudi, te raznih teoretskih sustava i budućih potencijalnih sintetičkih inteligencija.⁸

3.1. pristupi definiranju inteligencije

Osim neslaganja oko toga što točno pojam UOI predstavlja, jedna od nepremostivih prepreka cijelom području umjetne inteligencije predstavlja nedostatak jedinstvene koherentne definicije same inteligencije. Iako postoji nekakvo intuitivno shvaćanje što taj izraz podrazumijeva, nažalost samu inteligenciju nemoguće je definirati kao što je često slučaj s pojmovima u prirodnim jezicima. Unatoč tome, do danas je razvijeno nekoliko specifičnih pristupa u pokušaju rješavanja ovog problema.

3.1.1. Psihološki pristup

Jedan od osnovnih i najstarijih pristupa je psihološka mjera inteligencije. Među najranijim teorijama ovog područja izdvaja se takozvani g faktor kojeg je 1904. godine formulirao Charles Spearman. On je vjerovao kako je ovaj faktor biološki uvjetovan te da predstavlja ukupnu razinu intelektualne vještine pojedinca. Drugi bitan događaj nastupio je 1916. s definiranjem poznatog kvocijenta inteligencije (*intelligence quotient - IQ*), koji se izračunava dijeljenjem mentalne dobi ispitanika s njegovom fizičkom ili kronološkom dobi. U narednim godinama psiholozi su počeli sumnjati u pojam jedinstvene inteligencije. Postojala su dva primarna razloga. Prvenstveno, dokazano je da iako su performanse pojedinca u različitim domenama znanja u određenoj korelaciji, nije neobično da je njegova razina vještina u jednoj domeni znatno viša ili niža nego u drugoj. Ovo je poznato kao intra-individualna varijabilnost. Drugo, dvije osobe sa sličnim ukupnim performansama mogu se značajno razlikovati u specifičnim domenama znanja, što je poznato kao inter-individualna varijabilnost. Opažanja ovih fenomena rezultirala su stvaranjem cijelog niza alternativnih teorija, definicija i pristupa koji dijele ideju da je inteligencija višeslojna i varijabilna unutar i između pojedinaca. Među njima posebno se ističe teorija višestrukih inteligencija Howarda Gardnera, prema kojoj postoji osam različitih oblika ili vrsta inteligencije: 1) jezična, 2) logičko-matematička, 3) glazbena, 4) tjelesno-kinestetska, 5) prostorna, 6) interpersonalna, 7)

⁸ Ibid.

intrapersonalna i 8) naturalistička. Njegova teorija sugerira da intelektualnu vještinu svakog pojedinca predstavlja profil inteligencije, odnosno jedinstveni mozaik ili kombinacija razina vještina unutar osam oblika inteligencije.⁹

Današnje prevladavajuće stajalište je da ljudska inteligencija nije toliko općenita. Štoviše, ogromna količina naše inteligencije usmjerena je na situacije koje su se dogodile u našem evolucijskom iskustvu, kao što su socijalne interakcije, obrada vida, kontrola pokreta i tako dalje. Na primjer, većina ljudi je loša u apstraktnim procjenama vjerojatnosti, ali kada su te iste procjene stavljene u kontekst poznatih društvenih situacija ljudska točnost postaje znatno veća. Iako naša inteligencija je općenita u načelu, kako bi uspješno rješavali određene probleme moramo se koristiti sporim metodama poput matematike ili programiranja. S druge strane, znatno smo učinkovitiji u rješavanju problema koji zahtijevaju korištenje naših ugrađenih specijaliziranih neuronskih sklopova za obradu vida, zvuka, jezika, socijalnih interakcija itd. Gardnerov zaključak je da različiti ljudi imaju posebno učinkovite sklopove za različite specijalnosti. Također, jedan od zanimljivijih pristupa je onaj psihologa Roberta Sternberga koji razlikuje tri aspekta inteligencije, komponentni, kontekstualni i iskustveni. Komponentni se odnosi na specifične vještine koje ljudi posjeduju te koje ih čine inteligentnima. Kontekstualni predstavlja sposobnost uma da razumije i djeluje unutar konteksta te mogućnost odabira i mijenjanja tog konteksta, a iskustveni aspekt odnosi se na sposobnost uma da uči i prilagođava se kroz iskustvo.

Navedene teorije podrazumijevaju da UOI sustavi dizajnirani i izrađeni prema psihološkom pristupu moraju posjedovati sljedeće:

- sposobnost rješavanja općih problema koji nisu ograničeni na jedno područje, u istom smislu u kojem ljudi to mogu
- sposobnost rješavanja problema u pojedinim područjima i određenim kontekstima s posebnom učinkovitošću
- sposobnost da zajedno koriste svoju više općenitu i više specijaliziranu inteligenciju na jedinstven način
- sposobnost da uče iz svoje okoline, drugih inteligentnih sustava i učitelja
- sposobnost da postanu bolji u rješavanju novih vrsta problema kako stječu iskustvo¹⁰

⁹ Ibid.

¹⁰ Pennachin, C.; Goertzel, B. Contemporary Approaches to Artificial General Intelligence. AGIRI - Artificial General Intelligence Research Institute, 1405 Bernerd Place, Rockville, MD 20851, SAD, 2007.

Bitno je naglasiti kako se ove karakteristike u velikoj mjeri baziraju na ljudskom konceptu inteligencije te su kao takve znatno antropomorfne. Ovakav princip implementacije specijalizirane višestruke inteligencije nije opće prihvaćen unutar UI područja. Štoviše, neki znanstvenici inzistiraju na mogućnosti postojanja UOI sustava koji bi bio tako dobar u općenitom aspektu inteligencije da mu specijalizirane komponente ne bi bile potrebne. Jedan veoma bitan aspekt ovakvog tipa inteligencije je da se ona može postići samo kroz sustav koji je sposoban za učenje, a pogotovo za samostalno i inkrementalno učenje. Takav sustav bi trebao posjedovati mogućnost komuniciranja sa svojom okolinom, kao i s drugim subjektima u njoj, te učenja iz tih interakcija. On bi također trebao biti sposoban graditi na svojim dosadašnjim iskustvima i vještinama koje je naučio kako bi mogao obavljati kompleksnije radnje i postizati kompleksnije ciljeve.¹¹ Ako pogledamo dosadašnje uspješne programe iz uske UI domene evidentno je kako se velika većina njih odnosila upravo na visoko specijalizirane sposobnosti kao što su igranje šaha, dokazivanje teorema, prepoznavanje govora itd. Iako su određena istraživanja provedena na području kompletno općenitih UOI algoritama koji nisu orijentirani na specifične domene, ona do danas nisu rezultirala s praktičnim rješenjima.

3.1.2. Pragmatični pristup

Jedan od glavnih predvodnika ovakvog pristupa je UI istraživač Nils Nillson, a njegova razmišljanja su najbolje sažeta u sljedećoj rečenici: „Za postizanje stvarne ljudske razine u umjetnoj inteligenciji trebalo bi biti moguće automatizirati većinu poslova za koje su ljudi plaćeni. Umjesto rada na cilju automatizacije izgradnjom specijaliziranih sustava, zalažem se za izgradnju sustava opće namjene koji su sposobni učiti te obavljati bilo koji od tisuću poslova za koje su ljudi plaćeni.“¹² Suština ovakvog pogleda je da neki stroj, program ili arhitektura posjeduju ljudsku inteligenciju ako učine ljude zastarjelima ili nepotrebnima u određenom području. Pretpostavka kojom se Nillson vodi je da su ljudi općenito inteligentni sustav kojega želimo emulirati u računalnom okruženju, tako da je najbolji način identificiranja UOI sustava usporedba s ljudskim sposobnostima. Ovakav pristup je u nekoj mjeri podosta sličan klasičnom Turingovom testu u kojem se ispitivač upušta u razgovor na prirodnom jeziku s jednim čovjekom i jednim računalom koji je dizajniran kako bi imitirao

URL: http://www.springer.com/cda/content/document/cda_downloadaddocument/9783540237334-c1.pdf?SGWID=0-0-45-330089-p43950079 (20.8.2016.)

¹¹ Ibid.

¹² Nilsson, N. J. Human-Level Artificial Intelligence? Be Serious!, u: *AI Magazine*, vol. 26, br. 4, 2005. Str. 68-75.
URL: <http://www.aaai.org/ojs/index.php/aimagazine/article/view/1850/1748> (20.8.2016.)

ljudske odgovore. Prilikom testiranja svi sudionici su odvojeni jedni od drugih, a ako sudac ne može pouzdano razlikovati stroj od čovjeka, smatra se kako je stroj prošao test. Bitno je naglasiti kako sam Turing nije nikada tvrdio da njegov test mjeri inteligenciju ili neku drugu ljudsku osobinu, već sposobnost stroja da se ponaša inteligentno. U konačnici, Turingova ideja, kao i prethodno opisani pragmatični pristup, je eksplicitno antropomorfna. Naravno brojni teoretičari ne smatraju kako je njegov test dovoljan kriterij za UOI, a jedan od poznatijih primjera je argument Kineske sobe o kojem će više biti rečeno poslije. No jedna bitna razlika leži u tome što je Turing stavljao fokus na emulaciju ljudi, odnosno njihovog ponašanja. S druge strane, Nillsona ne zanima može li stroj zavarati ljude već može li obavljati korisne i praktične stvari koje ljudi mogu raditi.¹³

3.1.3. Pristup teorije kontrole

Teorija kontrole je interdisciplinarna grana inženjerstva i matematike koja se bavi ponašanjem dinamičkih sustava s ulazima te kako se njihovo ponašanje modificira na temelju povratnih informacija.¹⁴ Laički rečeno, fokus ove grane je dizajn strojeva koji reagiraju na unutarnje i vanjske podražaje te na temelju toga mijenjaju svoje ponašanje. Navedeno područje nudi nekoliko zanimljivih definicija inteligencije. Jedna od njih navodi kako se stroj može smatrati inteligentnim ako na temelju podataka koji su mu pruženi tijekom „treninga“, zajedno s informacijama o željenoj radnji, poduzima ispravne akcije na skupu drugih podataka unutar istog konteksta.¹⁵ Iako možda odmah nije očito, navedena definicija ne nudi mnogo po pitanju dizajna UOI sustava. Naime, ovakav pristup je karakterističan za većinu današnjih UI programa koji su itekako sposobni generalizirati svoje znanje, ali samo unutar ograničenog konteksta. No jedna druga teorija unutar ovog područja je puno zanimljivija. Ona naime tvrdi kako je inteligencija sposobnost ponašanja na odgovarajući način u nepredvidivim okolnostima. Iako je ova definicija ponešto nejasna, ona isključuje postojeće UI sustave zato što oni djeluju unutar striktno predvidivih okolina. Štoviše, navedeni uvjet o nepredvidivosti u određenoj mjeri aludira na iskustveni i kontekstualni aspekt psihološke analize inteligencije Roberta Sternberga. Jedna od nadogradnji ovakvog argumenta navodi kako je inteligencija

¹³ Goertzel, B. Artificial General Intelligence: Concept, State of the Art, and Future Prospects, u: *Journal of Artificial General Intelligence*, vol. 5, br. 1, 15. ožujka 2014. Str. 1-48.

URL: <https://www.degruyter.com/downloadpdf/j/jagi.2014.5.issue-1/jagi-2014-0001/jagi-2014-0001.xml> (20.8.2016.)

¹⁴ Control theory. Wikipedia. URL: https://en.wikipedia.org/wiki/Control_theory (20.8.2016.)

¹⁵ Pennachin, C.; Goertzel, B. Contemporary Approaches to Artificial General Intelligence. AGIRI - Artificial General Intelligence Research Institute, 1405 Bernerd Place, Rockville, MD 20851, SAD, 2007.

URL: http://www.springer.com/cda/content/document/cda_downloadaddocument/9783540237334-c1.pdf?SGWID=0-0-45-330089-p43950079 (20.8.2016.)

ostvarivanje kompleksnih ciljeva u kompleksnom okruženju. Iako su obje definicije korisne, ono što je očito je veliki stupanj relativnosti i subjektivnosti pojmova koje koriste. Izrazi poput kompleksnih ciljeva i okruženja, odgovarajućeg ponašanja i nepredvidivih okolnosti nisu dovoljno jasno definirani te se mogu radikalno razlikovati od sustava do sustava. U konačnici, ovakav pristup ne određuje nikakva posebna unutarnja svojstva inteligentnih strojeva. Suštinski, on naglašava što inteligencija radi, a ne kako. Međutim, isto tako ostavlja mogućnost da jedno implicira drugo u smislu da mogu postojati određene strukture i procesi koji su potrebni aspekti nekog inteligentnog sustava.¹⁶

3.1.4. Efikasna inteligencija

Zanimljivu definiciju inteligencije ponudio je Pei Wang, jedan od vodećih znanstvenika područja umjetne inteligencije. Wang tvrdi kako je inteligencija sposobnost sustava da se prilagodi svojoj okolini pri čemu raspolaže s nedovoljno znanja i resursa.¹⁷ U ovome kontekstu, okolina može biti fizički svijet ili neki drugi informacijski sustav dok se pod pojmom prilagodbe podrazumijeva da sustav odlučuje o svojim akcijama na temelju prijašnjih iskustava kako bi bolje postizao svoje ciljeve. Osnovni zaključak ove definicije je da sustav mora funkcionirati pod takozvanom pretpostavkom nedovoljnog znanja i resursa (*Assumption of Insufficient Knowledge and Resources - AIKR*). Konkretno, to znači da sustav mora posjedovati sljedeće karakteristike::

- Konačan: računalna snaga, u smislu brzine procesora i prostora za pohranu, je ograničena,
- Stvarno vrijeme: sustav se mora nositi s problemima koji se mogu pojaviti u bilo kojem trenutku te koji zahtijevaju hitnu reakciju,
- Otvoren: sustav se mora suočiti s ulaznim podacima i problemima bilo kojeg sadržaja, dok god ih se može izraziti u obliku koji je prepoznatljiv sustavu.¹⁸

Za razliku od prethodnog pristupa ovakva definicija nije samo bihevioralna nego se bavi unutarnjim aspektima UOI sustava čija se inteligencija procjenjuje. No možda najveća distinkcija leži u naglasku na ograničenju računalne snage sustava. Ako za primjer uzmemo

¹⁶ Ibid.

¹⁷ Wang, P. The Logic of Intelligence. Department of Computer and Information Sciences, Temple University Philadelphia, PA 19122, SAD, 2006.

URL: http://www.cogsci.indiana.edu/farg/peiwang/PUBLICATION/wang.logic_intelligence.pdf (20.8.2016.)

¹⁸ Wang, P. Non-Axiomatic Logic: A Model of Intelligent Reasoning. Department of Computer and Information Sciences, Temple University Philadelphia, PA 19122, SAD, 2013.

URL: <http://cis-linux1.temple.edu/~pwang/Publication/Diagnosis.pdf> (20.8.2016.)

neki hipotetski UI algoritam koji pretpostavlja beskonačne računalne resurse, prema navedenoj definiciji takav algoritam ne može posjedovati inteligenciju. Štoviše, Wang smatra kako bi takav program samo emulirao inteligenciju pomoću jednostavnih mehanizama za pretraživanje.¹⁹ Ono što je također implicirano ovim pristupom je nedostatnost tradicionalne matematičke logike za izgradnju UI programa koji funkcioniraju pod *AIKR* pretpostavkom. Naime, ovakve metode poput logike prvog reda namijenjene su za zatvorene aksiomske sustave koji djeluju u idealiziranim situacijama gdje sustav ima dovoljno znanja i resursa u odnosu na probleme koje treba riješiti. Za razliku od takvog pristupa, inteligentni agent koji djeluje pod *AIKR* pretpostavkom ne posjeduje početni skup aksioma koji garantiraju istinitost njegovih zaključaka. Ovakav sustav temelji svoje zaključke na prijašnjim iskustvima koja se koriste kako bi se predvidjela budućnost pri čemu njegova predviđanja itekako mogu biti kriva.²⁰

3.1.5. Utjelovljeni pristup

Radikalno drugačiji pristup formuliran je u takozvanoj utjelovljenoj perspektivi. U osnovi, zagovornici ovog pristupa smatraju kako je inteligencija nešto što fizička tijela rade u fizičkom okruženju. Prema ovakvom stajalištu, inteligentni agenti uvijek se pokoravaju fizičkim i društvenim pravilima svoje okoline te iskorištavaju ta pravila kako bi proizveli raznolika ponašanja.²¹ Kao osnovni argument u prilog potrebi za fizičkim tijelom navodi se primjer evolucije. Točnije, primjer ljudske inteligencije koja se posebno prilagodila zadatku kontroliranja ljudskog tijela u kompleksnom okruženju u uvjetima oskudnih energetske resursa i fizičkih ograničenja. Kao najveći zagovornik ovog pristupa istaknuo se Rodney Brooks sa svojim istraživanjima na području robotike. Naime, Brooks se proslavio tijekom 60-ih i 70-ih godina prošlog stoljeća kada je uspješno konstruirao nekoliko robota bez korištenja tradicionalnih simboličkih sustava, nešto što se tada smatralo nemogućim. Njegova istraživanja u velikoj su mjeri pridonijela formulaciji Moravecova paradoksa. Svoje stavove Brooks je potvrdio u knjizi *Elephants Don't Play Chess* gdje je naveo kako je svijet svoj

¹⁹ Pennachin, C.; Goertzel, B. Contemporary Approaches to Artificial General Intelligence. AGIRI - Artificial General Intelligence Research Institute, 1405 Bernerd Place, Rockville, MD 20851, SAD, 2007.
URL: http://www.springer.com/cda/content/document/cda_downloadaddocument/9783540237334-c1.pdf?SGWID=0-0-45-330089-p43950079 (20.8.2016.)

²⁰ Wang, P. The Logic of Intelligence. Department of Computer and Information Sciences, Temple University Philadelphia, PA 19122, SAD, 2006.
URL: http://www.cogsci.indiana.edu/farg/peiwang/PUBLICATION/wang.logic_intelligence.pdf (20.8.2016.)

²¹ Goertzel, B. Artificial General Intelligence: Concept, State of the Art, and Future Prospects, u: *Journal of Artificial General Intelligence*, vol. 5, br. 1, 15. ožujka 2014. Str. 1-48.
URL: <https://www.degruyter.com/downloadpdf/j/jagi.2014.5.issue-1/jagi-2014-0001/jagi-2014-0001.xml> (20.8.2016.)

najbolji model te sadrži svaki detalj kojeg je potrebno poznavati.²² Utjelovljeni pristup u konačnici je rezultirao nastankom zasebnog područja unutar umjetne inteligencije, nazvan *Nouvelle UI*. No brojni istraživači do danas se ne slažu s ovakvim pretpostavkama. Jedan od većih izvora prijepora je konstatacija kako se inteligencija može pripisati samo fizičkim sustavima čije se ponašanje može promatrati kroz njihovu interakciju s okolinom. Protivnici ovakvog stajališta, među kojima se ističe Wang, smatraju kako je svaki softver s određenom vrstom korisničkog sučelja u interakciji s fizičkim svijetom putem neke vrste tijela.²³

3.2. Tipologija UOI sustava

Ako se prihvati pretpostavka da UOI pokušava kreirati sustave koji posjeduju inteligenciju sličnu ljudskoj, onda je potrebno specificirati u kojem točno području te sličnosti leže. Glavni razlog ovakve klasifikacije je nerealnost očekivanja kako će takvi sustavi nalikovati ljudima u svakome aspektu. Bez ulaženja u prevelike detalje moguće je navesti pet osnovnih kategorija koje se razlikuju po području sličnosti između UI programa i ljudi.

3.2.1. Struktura

Budući da se najpoznatiji primjer inteligencije nalazi unutar ljudskoga mozga, jedna od prirodnijih pretpostavki je da se UI može postići izgradnjom konstrukcije mozga koja će se sastojati od ogromnog skupa procesorskih jedinica nalik neuronima. Ovaj pristup je razrađen u brojnim formama, od kojih su najpoznatije ideje konekcionizma i umjetnih neuronskih mreža. Naravno, zbog same kompleksnosti mozga, ali i njegove različitosti od računalnog hardvera, ovakvi projekti služe samo kao aproksimacija njegove strukture. Unatoč tome, danas postoje brojni poklonici ovog pristupa koji smatraju kako će napretkom našeg znanja i tehnologije ovakvi projekti jednoga dana omogućiti izgradnju potpunih UI programa. Glavna prednost strukturne UI leži u njezinom proučavanju ljudskog mozga. Konkretno, ovaj pristup znatno doprinosi boljem razumijevanju raznih kognitivnih aktivnosti. No ako konačni cilj projekta izgradnje UI sustava leži u nekoj od ostalih kategorija, dupliciranje strukture mozga

²² Rodney A. Brooks. Elephants Don't Play Chess, u: *Robotics and Autonomous Systems*, vol. 6, br. 1-2, lipanj 1990. Str. 3-15. URL: <http://people.csail.mit.edu/brooks/papers/elephants.pdf> (20.8.2016.)

²³ Wang, P. Embodiment: Does a laptop have a body? Department of Computer and Information Sciences, Temple University, Philadelphia, PA 19122, SAD, 2009. URL: http://agi-conference.org/2009/papers/paper_5.pdf (20.8.2016.)

je poprilično neefikasan pristup zato što je on formiran pod biološkim i evolucijskim ograničenjima koja su uglavnom nebitna za računala.²⁴

3.2.2. Ponašanje

Mnogi vjeruju kako je inteligencija karakteristika uma, a ne mozga, te samim time kako je ponašanje sustava puno bolji kriterij za njegovu evaluaciju. Vjerojatno najpoznatiji primjer ove kategorije je već spomenuti Turingov test, no drugi dobar primjer je područje chatbotova gdje se inteligencija sustava ocjenjuje prema tome koliko on govori kao čovjek. Prema ovom pristupu UI se poistovjećuje s kognitivnim modeliranjem gdje se računalno proizvedeni rezultati ocjenjuju usporedbom s psihološkim rezultatima proizvedenim od strane ljudskih subjekata. U ovome kontekstu sustavi se tretiraju kao crne „kutije“ čije unutarnje strukture i stanja nisu bitni. Prednost bihevioralne UI leži u njezinom doprinosu proučavanju ljudske psihologije, no s druge strane ona često nije najbolji način za rješavanje praktičnih problema ili implementaciju kognitivnih funkcija.²⁵

3.2.3. Sposobnost

Za one čiji interes u UI leži u njezinim potencijalnim praktičnim primjenama, inteligenciju sustava treba evaluirati njegovom sposobnošću rješavanja teških problema. Dijelomično vođena takvim motivima, najraniji praktični problemi kojima se UI bavila bili su tipične intelektualne aktivnosti kao dokazivanje teorema i igranje raznih igara. Isto tako, veliki broj aplikacijsko orijentiranih UI projekata su ekspertni sustavi u različitim domenama. Generalni zaključak je da ako računalo može riješiti problem koji je rezerviran za stručnjake, onda računalo mora biti inteligentno. Najpoznatiji primjeri bili bi *Deep Blue*, šahovski sustav koji je porazio svjetskog prvaka, ili *AlphaGo*, Googleov sustav koji je uspio poraziti nekoliko prvaka u tradicionalnoj kineskoj igri Go. Prema ovome stajalištu UOI programi bili bi u stanju rješavati sve veće i kompleksnije probleme, a jedan primjer bio bi prethodno spomenuti Nilssonov test zapošljavanja. Među postojećim UOI projektima izdvojio bi se Cyc koji kodira ogromne količine zdravorazumskog znanja kako bi postigao ljudsku razinu rješavanja problema. Iz očitih razloga ovakva UI kroz rješavanje praktičnih problema znatno doprinosi različitim područjima primjene. Međutim, zbog čestog nedostatka općenitosti tih rješenja ovaj

²⁴ Wang, P. What Do You Mean by “AI”? u: *Proceedings of the 2008 conference on Artificial General Intelligence*, Memphis, Tennessee, ožujak 2008., str. 362-373.

URL:http://cis-linux1.temple.edu/~wangp/Publication/AI_Definitions.pdf (20.8.2016.)

²⁵ Ibid.

pristup relativno je ograničen u proučavanju problema koji su van opsega veoma specifičnih domena.²⁶

3.2.4. Funkcija

Većinu istraživača UI čine računalni znanstvenici i inženjeri. U skladu s time oni preferiraju predstavljati sposobnost agenta kao neku funkciju koja mapira ulazne podatke u izlazne akcije. Uostalom, ovo i je standardan način definiranja računalnog programa. Sistematski opis ovoga pristupa pružio je Marr koji kaže kako rezultat u umjetnoj inteligenciji podrazumijeva identifikaciju određenog problema obrade informacija, formulaciju računalne teorije, izradu algoritma koji ju implementira te praktičnu demonstraciju uspješnosti algoritma. Na temelju takvih mišljenja mnogi smatraju kako se područje umjetne inteligencije sastoji od odvojenih kognitivnih funkcija kao što su pretraživanje, zaključivanje, planiranje, učenje, rješavanje problema, donošenje odluka, komuniciranje itd., pri čemu svaki problem ima svoje različite računalne formulacije i algoritamske implementacije. U kontekstu UOI ova definicija podrazumijeva da sustav mora imati mnoge kognitivne funkcije koje rade zajedno. Primjeri bi uključivali *LIDA* i *Novamente* kognitivne arhitekture. Funkcijska UI prvenstveno doprinosi razvoju računalne znanosti kroz izradu raznih specifičnih softverskih rješenja. Međutim, zbog toga što se ljudske kognitivne funkcije generaliziraju na rasponu mnogih konkretnih problema, ovakav specijalizirani pristup nije nužno i najbolji.²⁷

3.2.5. Načelo

Zadnji pristup mogao bi se okarakterizirati kao općenito znanstveni. Naime, vođeni idejom jednostavnih i jedinstvenih objašnjenja kompliciranih pojava, pobornici ovakvog pristupa pokušavaju identificirati fundamentalno načelo po kojem bi se ljudska inteligencija mogla objasniti i reproducirati u računalima na općoj razini. Budući da je ovakav pristup preopćenit, predloženi su određeni kompromisi poput ograničene racionalnosti koja ograničava ono što agent može znati i činiti na temelju kognitivnih limitacija i raspoloživog vremena²⁸, ili pak ograničene optimalnosti koja određuje sposobnost agenta za generiranje maksimalno uspješnog ponašanja s obzirom na dostupne informacije i računalne resurse. Primjer bi bio Wangov predloženi neaksiomatski sustav za rasuđivanje (*Non-Axiomatic Reasoning System - NARS*) koji funkcionira pod pretpostavkom nedovoljnog znanja i resursa.

²⁶ Ibid.

²⁷ Ibid.

²⁸ Bounded rationality. Wikipedia. URL: https://en.wikipedia.org/wiki/Bounded_rationality (20.8.2016.)

Kao i kod prethodnog pristupa i ovdje je fokus na funkciji. Razlika je u tome što se ona ovdje naziva načelo kako bi se naglasilo da se ne radi o samo jednom problemu i njegovom rješenju već o cjelokupnom kontekstu agenta u različitim situacijama koji se bavi raznim vrstama problema. Ovakva načelna UI doprinosi proučavanju obrade podataka u raznim situacijama istražujući implikacije različitih pretpostavki. S obzirom na njezinu široku općenitost ona ne može objasniti sve detalje o ljudskom mozgu ili umu, niti pruža najbolji način rješavanja svakog praktičnog problema. Iako sustavi koji se temelje na ovom pristupu obično sadrže razne kognitivne funkcije, one nisu nužno odvojeni procesi sa svojim vlastitim računalnim formulacijama i algoritamskim implementacijama.²⁹

Za kraj je potrebno napomenuti kako su sve prethodne definicije validne te ne moraju nužno isključivati jedna drugu. No isto tako, svaka od njih odnosi se na opis ljudske inteligencije na različitim razinama apstrakcije. U skladu s time, svaka kategorija podrazumijeva različite ciljeve dizajna i izgradnje UOI sustava kao i različite kriterije evaluacije. Jedan od osnovnih problema današnjeg područja umjetne opće inteligencije je taj što većina projekata nema jasno određene ciljeve istraživanja zato što se istovremeno koriste različitim definicijama inteligencije. U konačnici, takav pristup uglavnom rezultira s nekonzistentnim kriterijima projektiranja i evaluacije sustava.

²⁹ Wang, P. What Do You Mean by "AI"?, u: *Proceedings of the 2008 conference on Artificial General Intelligence*, Memphis, Tennessee, ožujak 2008., str. 362-373.
URL: http://cis-linux1.temple.edu/~wangp/Publication/AI_Definitions.pdf (20.8.2016.)

4. IZRADA UOI SUSTAVA

Kao što je prethodno spomenuto, cjelokupno područje umjetne opće inteligencije fragmentirano je na veliki skup međusobno konfliktnih područja. Svako od njih djeluje pod određenim pretpostavkama te implementira različite metode dizajna i izrade inteligentnih sustava. Sama količina različitih pristupa prevelika je kako bi se kvalitetno prikazala u ovome radu. No unatoč različitim praktičnim metodama, postoje određene konceptualne zakonitosti jednako kao i u području tradicionalne umjetne inteligencije. Ono što se pod time podrazumijeva je famozna dihotomija između simboličkog i podsimboličkog pristupa. U narednim poglavljima ukratko će biti prikazana oba pristupa te neke njihove najvažnije karakteristike.

4.1. Simbolički pristup

Jedan od osnovnih koncepata unutar ovog područja predstavlja takozvana simbolička UI. Štoviše, većina ambicioznih UOI orijentiranih projekata poduzetih do danas bili su u simboličnoj UI paradigmi. Sama ideja vuče svoje korijene iz razmišljanja mnogih filozofa poput Thomasa Hobbesa, Davida Humea i Immanuela Kanta koji su smatrali kako se svako životno iskustvo može svesti pod formalna pravila. Osnovne postavke simbolične UI artikulirane su 1976. godine pod nazivom hipoteza sustava fizičkih simbola (*physical symbol system hypothesis*) od strane Allena Newella i Herberta A. Simona. Osnovni elementi ovog koncepta su skupine entiteta, nazvani simboli, koji su fizički uzorci koji se mogu pojaviti kao sastavnica druge vrste entiteta pod nazivom izrazi, odnosno struktura simbola. Dakle struktura simbola sastoji se od niza instanci, ili tokena, simbola povezanih u nekom fizičkom smislu. U bilo kojem trenutku sustav će sadržavati skup ovakvih struktura. Osim navedenih struktura, sustav sadrži skupove procesa koji funkcioniraju na danim izrazima kako bi proizveli druge izraze. To su procesi kreiranja, modificiranja, reproduciranja i uništavanja.³⁰ Dva pojma su središnja u ovoj strukturi izraza, simbola i objekata, a to su označivanje i interpretacija. Označivanje je ideja da se izrazi mogu odnositi na nešto drugo. Odnosno, simbol je samo simbol osim ako postoji neki smisao u kojemu on ima funkciju reference. Potreba za ovime postoji jer naglasak nije na stroju koji samo računa već na stroju koji se „ponaša“ u odnosu na svijet koji ga okružuje. Simbolički izraz označava neki objekt ako je ponašanje sustava prema vanjskom svijetu ovisno o objektu na koji se upućuje. Ideja je da se inteligentno ponašanje

³⁰ Newell, A.; Simon, H. A. Computer Science as Empirical Inquiry: Symbols and Search, u: *Communications of the ACM*, vol. 19, br. 3, ožujak 1976. Str. 113-126.

URL: <http://www.cs.utexas.edu/~kuipers/readings/Newell+Simon-cacm-76.pdf> (20.8.2016.)

realizira u odnosu na neki vanjski svijet. Dakle, promišljanje o tome kako postupati u ovome svijetu zahtijeva da sustav strukture simbola u nekom smislu predstavlja aspekte tog vanjskog svijeta. Interpretacija je dodatno proširenje ove ideje reprezentacijske moći strukture simbola. Ovdje je ideja da se struktura simbola ili izraz može odnositi na računalni proces kojeg sustav može interpretirati i provoditi. Ovakva ideja poznata je pod krinkom računalnog programa. Statički gledano, računalni program je samo zbirka izraza. Ali, ako je program valjan, onda se taj skup izraza može pročitati i pretvoriti u proces koji obavlja upute programa. Nekakav generalni zaključak ove teorije bio bi da umovi postoje kako bi manipulirali simbolima koji predstavljaju aspekte svijeta ili njih samih. Štoviše, Newell i Simon bili su toliko uvjereni u univerzalnost ovakve teorije da su tvrdili kako fizički sustav simbola posjeduje potrebna i dovoljna sredstva za ostvarivanje opće inteligentnog djelovanja.³¹

Ovaj pristup primijenjen je na ogromnom broju teoretskih i praktičnih problema što je rezultiralo izradom brojnih raznolikih sustava. Jedan od prvih takvih programa usmjeren na opću inteligenciju bio je *General Problem Solver* iz 1957. godine. Newell i Simon dizajnirali su ga kako bi pomoću heuristike rješavao različite probleme poput hanojskih tornjeva ili kript-aritmetike. GPS je u osnovi funkcionirao razbijanjem glavnog cilja na skup sve manjih podciljeva koji su se onda mogli direktno obraditi jednostavnom heuristikom.³² Nakon demonstracije uspješnosti ovakvih programa simbolički pristup implementiran je na nizu različitih problema kao što su procesi zaključivanja, rješavanje problema, komunikacija na prirodnom jeziku, robotika itd. Među poznatijim programima tog razdoblja izdvajaju se *Geometry Theorem Prover* iz 1958. koji je rješavao geometrijske zadatke, te STUDENT i SHRDLU s kraja 70-ih koji su se bavili obradom prirodnog jezika. No dva područja koja su od presudne važnosti za opću inteligenciju su prikaz znanja i učenje.

4.1.1. Prikaz znanja

Većina tradicionalnih računalnih sustava oslanja se na procese simboličke reprezentacije i manipulacije za prikaz i obradu znanja. Točnije, znanje se prikazuje u striktno formalnom obliku, a možda najpoznatiji primjer primjene ovakvog pristupa su ekspertni sustavi sa svojim bazama znanja. Ovdje se radi o fundamentalnoj ideji nastaloj u ranim danima umjetne inteligencije te je kao takva odigrala centralnu ulogu u njezinom razvoju.

³¹ Ibid.

³² Pennachin, C.; Goertzel, B. Contemporary Approaches to Artificial General Intelligence. AGIRI - Artificial General Intelligence Research Institute, 1405 Bernerd Place, Rockville, MD 20851, SAD, 2007.
URL: http://www.springer.com/cda/content/document/cda_downloaddocument/9783540237334-c1.pdf?SGWID=0-0-45-330089-p43950079 (20.8.2016.)

Kroz nekoliko desetljeća razvijen je cijeli niz različitih oblika simboličkog prikaza znanja, a većina njih koristi se u kombinaciji s nekim algoritmom za pretraživanje koji omogućuje proces zaključivanja. Među najranije oblike ubraja se takozvano zaključivanje na temelju pravila (*rule-based reasoning*) u kojemu se koriste diskretna pravila za usmjeravanje pretraživanja, odnosno donošenje zaključaka. Ovakav formalizam sastoji se od uvjeta koji definiraju primjenjivost pravila i zaključaka koji definiraju radnje ili rezultate. On je također modularan u smislu da se svako pravilo može dodavati ili brisati bez utjecaja na ostatak sustava.³³ Posebno poznata podvrsta ovakvog oblika su pravila produkcije (*production rules*) koja su nastala na temelju psiholoških teorija iz 60-ih i 70-ih godina prošlog stoljeća. Ona pružaju mehanizam potreban za izvršenje produkcija kako bi se postigao neki cilj sustava, a sastoje se od preduvjeta (u obliku *IF* izjava) te radnji (u obliku *THEN* izjava). Produkcijski sustavi sastoje se od:

1. Baze produkcijskih pravila
2. Radne memorije koja sadržava podatke o trenutnom stanju sustava i rezultate radnji
3. Kontrolne strukture koja koordinira radom sustava³⁴

Unutar sustava temeljenih na pravilima alternativu predstavlja formalna simbolička logika, a kao njezin glavni zagovornik ističe se John McCarthy. Radi se relativno jednostavnim formalno definiranim jezicima koji se koriste za formulaciju pravila. Najčešći oblik je predikatni račun prvog reda. Ovaj pristup odlikuje zaključivanje na temelju formalno definiranih načina koji garantiraju potpunost i točnost zaključaka kao i mogućnost izvođenja korištenjem raznih algoritama. S druge strane, formalna logika kao i većina produkcijskih sustava je poprilično restriktivna u smislu da je potrebno imati sve uvjete precizno definirane kako bi se proces zaključivanja mogao odvijati.

Relativno drugačiji pristup unutar simboličke UI odnosi se na prikaz strukture znanja te sadrži nekoliko oblika. Prvi primjer bio bi koncept okvira (*frames*) koje je prvi predložio Marvin Minsky. Okviri predstavljaju različite objekte u vidu njihovih atributa od kojih svaki ima svoj naziv i vrijednost. Organiziranjem znanja u okvire moguće je pristupiti svim relevantnim informacijama na jednome mjestu. Drugi primjer čine semantičke mreže (*semantic network*), a sastoje se od čvorova koji predstavljaju neki pojam te označenih

³³ Sun, R. Artificial Intelligence: Connectionist and Symbolic Approaches, u: *International Encyclopedia of the Social and Behavioral Sciences*, uredili Smelser, N. J. i Baltes, P. B. Pergamon/Elsevier, Oxford, 2001. Str. 783-789. URL: <http://www.cogsci.rpi.edu/~rsun/sun.encyc01.pdf> (20.8.2016.)

³⁴ Production system (computer science). Wikipedia.

URL: [https://en.wikipedia.org/wiki/Production_system_\(computer_science\)](https://en.wikipedia.org/wiki/Production_system_(computer_science)) (20.8.2016.)

poveznica koje reprezentiraju relacije između tih pojmova. Ovakve mreže omogućuju efikasan pristup znanju kroz organizaciju taksonomskih hijerarhija.³⁵ I treći primjer uključivao bi pojam skripti (*scripts*). Njihov začetnik Roger Schank smatrao je kako se korištenjem skripti mogu reprezentirati sekvence događaja u svakodnevnim situacijama. Budući da se takve situacije uglavnom sastoje od fiksnih sekvenci, njih je pomoću skripti moguće efikasno prepoznati i prikazati. Glavni nedostatak ovakvih pristupa sa stajališta opće inteligencije leži u tome što je strukture potrebno unaprijed odrediti te ručno kodirati. One su također fiksirane te ih se ne može dinamički mijenjati, a mogu postati preglomazne kako bi u punoj mjeri obuhvatile kompleksne situacije u stvarnom svijetu.³⁶

Zbog navedenih karakteristika veliko ograničenje ovakvih sustava, a pogotovo onih baziranih na logici, leži u nemogućnosti prikazivanja ograničenog znanja. Točnije, ovakvi pristupi ne mogu funkcionirati s parcijalnim ili nepotpunim informacijama. Za potrebe ovakvih situacija razvijeno je nekoliko posebnih pristupa unutar simboličke paradigme. Među njima se svakako ističe takozvana *fuzzy* logika. U suštini ovo je logički sistem koji pokušava formalizirati zaključivanje na temelju neodređenih i aproksimativnih informacija. Prema osnovnoj ideji ove logike za svaki koncept postoji skup objekata koji zadovoljavaju kriterije tog koncepta do određenog stupnja, a ti objekti se grupiraju u *fuzzy* podskupove. Ovi podskupovi sadrže elemente parova koji se sastoje od objekta i njegovog stupnja pripadnosti određenom konceptu.³⁷ Za razliku od standardne logike koja vrijednostima istine dodjeljuje vrijednosti 0 ili 1, kod *fuzzy* logike ovim varijablama moguće je pridružiti bilo koji realan broj između 0 i 1 što znači da je moguće konstruirati pojmove na temelju parcijalne istine. Drugi sustav osmišljen za ovakve potrebe je pristup vjerojatnosti (*probabilistic approach*) koji tretira uvjerenja kao probabilističke događaje te se koristi zakonima vjerojatnosti. Ovdje se često koristi Bayesova (*Bayesian*) vjerojatnost u obliku kvantitete koja reprezentira stanje znanja ili uvjerenja. Konkretno, Bayesov pristup prikazuje vjerojatnost kao stupanj uvjerenja da će se neki događaj zbiti pri čemu se određuje njegova prethodna vjerojatnost koja se onda ažurira na temelju novih podataka.³⁸ Problem s ovim pristupom sa stajališta opće inteligencije

³⁵ Semantic network. Wikipedia. URL: https://en.wikipedia.org/wiki/Semantic_network (20.8.2016.)

³⁶ Sun, R. Artificial Intelligence: Connectionist and Symbolic Approaches, u: *International Encyclopedia of the Social and Behavioral Sciences*, uredili Smelser, N. J. i Baltes, P. B. Pergamon/Elsevier, Oxford, 2001. Str. 783-789. URL: <http://www.cogsci.rpi.edu/~rsun/sun.encyc01.pdf> (20.8.2016.)

³⁷ Ibid.

³⁸ Bayesian probability. Wikipedia. URL: https://en.wikipedia.org/wiki/Bayesian_probability (20.8.2016.)

je da se proces ljudskog zaključivanja često ne podvrgava pretpostavkama i zakonima teorije vjerojatnosti, a u praksi često nije moguće izračunati njenu adekvatnu mjeru.³⁹

Kao najveća prepreka simboličkom pristupu u ovoj domeni nameće se problem zdravorazumskog znanja. Naime, ljudi u svakodnevnim situacijama koriste ogromne količine implicitnog i proceduralnog znanja. Ako će ovakav pristup ikada biti uspješan u emulaciji ljudske inteligencije onda će biti neophodno prevesti to znanje u eksplicitni oblik, odnosno deklarativnu formu u obliku brojnih činjenica i pravila.

4.1.2. Učenje

Iako su unutar ovog područja razvijene brojne sofisticirane i efikasne tehnike reprezentacije znanja, pristup učenju pokazao se kao veoma zahtjevan zadatak za simboličku umjetnu inteligenciju. Kao očiti razlog za to ističe se činjenica da je simbolička UI od svog začetka prirodno bila usmjerena na procese prikaza, a ne učenja. No unatoč tome, postignuti su određeni napretci na nekoliko polja. Kao jedan od efikasnijih pristupa pokazalo se genetičko programiranje, odnosno genetički algoritmi. Ovdje se radi o metaheurističkoj metodi optimiziranja koja je inspirirana evolucijskim procesima mutacije i selekcije. Kod genetičkog algoritma kao kriterij selekcije služi funkcija cilja koja predstavlja zadani problem. U odnosu na problem algoritam vrši operacije nad populacijom potencijalnih rješenja koja predstavljaju jedinke. Procesima selekcije i manipulacije nastaju nove populacije, a ovakva iteracija završava s najboljom jedinkom koja predstavlja optimalno rješenje.⁴⁰ U suštini ovo je robustan proces pretraživanja prostora rješenja.

Općenito gledajući, tijekom godina uloženi su veliki naponi kako bi se analitičko i induktivno učenje uvelo u područje simboličkih sustava. Određene induktivne metode, poput stabla odlučivanja ili već spomenutog genetičkog algoritma, pokazale su se efikasnim u određenim situacijama. Ovakvo učenje, jednostavno rečeno, nastoji izvući iz određenih činjenica ili primjera opća pravila koja odražavaju temeljne strukture neke domene. Induktivna metoda koristi pružene podatke tijekom treninga te pronalaženjem uzoraka među njima stvara opću hipotezu. Nedostatak ovog pristupa je što se hipoteze donose na temelju statističkog zaključivanja što znači da ova metoda zahtijeva velike skupove početnih podataka kako bi bila efikasna. Analitičko učenje pokušava riješiti ovaj problem uključivanjem

³⁹ Sun, R. Artificial Intelligence: Connectionist and Symbolic Approaches, u: *International Encyclopedia of the Social and Behavioral Sciences*, uredili Smelser, N. J. i Baltes, P. B. Pergamon/Elsevier, Oxford, 2001. Str. 783-789. URL: <http://www.cogsci.rpi.edu/~rsun/sun.encyc01.pdf> (20.8.2016.)

⁴⁰ Genetic algorithm. Wikipedia. URL: https://en.wikipedia.org/wiki/Genetic_algorithm (20.8.2016.)

pozadinskog znanja koje se naziva teorijom domene. Ova metoda koristi to znanje kako bi objasnila podatke pružene tijekom treninga te na temelju njih formulirala opću hipotezu korištenjem deduktivnog zaključivanja. Za razliku od prethodnog pristupa, hipoteze dobivene analitičkom metodom moraju odgovarati i podacima i teoriji domene.⁴¹ Najkorištenija metoda ovakve vrste je učenje na temelju objašnjenja (*explanation-based learning* - EBL) kao oblik strojnog učenja koji iskorištava vrlo jaku ili čak savršenu teoriju domene kako bi generalizirao ili stvarao koncepte na temelju dostupnih podataka. EBL se pokazao korisnim u smanjenju kompleksnosti prostora mogućih hipoteza. Naime, koristeći prethodno znanje na određenom skupu podataka moguće je identificirati značajke koje su irelevantne. Dobar primjer je učenje šaha. Teorija domene u ovome slučaju je potpuna jer ona podrazumijeva samo pravila igre. Na temelju tih pravila u načelu bi trebalo biti moguće izračunati najbolji potez u svakom trenutku. No u stvarnosti ovo postaje nemoguće zbog problema kombinatorne eksplozije. U takvim situacijama EBL metoda može na temelju primjera odrediti koje su relevantne značajke kako bi formirala opće pravilo.⁴² Jedan od glavnih područja primjene ove metode je obrada prirodnog jezika. Naravno osnovni preduvjet uspješnosti ovakvog pristupa je točnost pozadinske teorije.

U posljednje vrijeme razvijeni su brojni algoritmi unutar područja simboličke UI kako bi se povećala uspješnost različitih pristupa kao što su inkrementalno učenje ili učenje podrškom (*reinforcement learning*). Među poznatije trendove svakako spada implementacija ovakvih metoda unutar sofisticiranih kognitivnih arhitektura. Kako bi postigle robusniju sposobnost učenja ove arhitekture stavljaju naglasak na radnu memoriju koja po potrebi pristupa dugoročnom pamćenju te koriste centraliziranu kontrolu nad percepcijom, kognicijom i djelovanjem. Iako bi u načelu takve arhitekture trebale biti sasvim dovoljne zato što simbolički sustavi u teoriji posjeduju sposobnost univerzalne reprezentacije, u praksi se simboličke arhitekture često pokazuju manje učinkovitim u procesima učenja (prije svega proceduralnog učenja) i kreativnosti.⁴³ U narednom dijelu ukratko će biti opisano nekoliko poznatijih pokušaja ovog područja u izradi općenito inteligentnih programa.

⁴¹ Mitchell, T. M.; Thrun, S. B. Learning Analytically and Inductively, u: *Mind Matters: A Tribute to Alan Newell*, uredili Steier, D. M. i Mitchell, T. M. Erlbaum, 1996. Str. 85-111.

URL: https://www.ri.cmu.edu/pub_files/pub1/mitchell_tom_1996_1/mitchell_tom_1996_1.pdf (20.8.2016.)

⁴² Explanation-based learning. Wikipedia. URL: https://en.wikipedia.org/wiki/Explanation-based_learning (20.8.2016.)

⁴³ Goertzel, B. Artificial General Intelligence: Concept, State of the Art, and Future Prospects, u: *Journal of Artificial General Intelligence*, vol. 5, br. 1, 15. ožujka 2014. Str. 1-48.

URL: <https://www.degruyter.com/downloadpdf/j/jagi.2014.5.issue-1/jagi-2014-0001/jagi-2014-0001.xml> (20.8.2016.)

4.1.2.1. CYC

Vjerojatno najpoznatiji i najveći projekt izrade UOI programa korištenjem simboličkog pristupa je Cyc sustav. Izgradnju ove arhitekture započeo je Doug Lenat davne 1984. godine. Fundamentalna ideja projekta je kodiranje ogromne količine zdravorazumskog znanja, odnosno svakodnevnih činjenica u bazu znanja koristeći predikatni račun prvoga reda. Na ovoj bazi primjenjuju se logičke tehnike zaključivanja kako bi se odgovaralo na razna pitanja i stvaralo novo znanje. Dio filozofije na kojoj se Cyc temelji je da nakon što se akumulira dovoljna količina znanja, problem stvaranja opće inteligencije slične ljudskoj će postati mnogo lakši. Ovo je ujedno i osobno uvjerenje Lenata koji smatra kako je jedini način da strojevi razumiju ljudske koncepte taj da ih se nauči jedan po jedan. Baza znanja trenutačno sadrži preko milijun definiranih tvrdnji, pravila i ideja koje se formuliraju u programskom jeziku CycL korištenjem sintakse slične onoj u LISP-u, a novija nadogradnja sustava uključuje spajanje zasebnog modula za obradu prirodnog jezika.⁴⁴

4.1.2.2. SOAR

Soar (*State, Operator and Result*) predstavlja klasičan primjer ekspertne kognitivne arhitekture temeljene na pravilima koja je osmišljena za modeliranje opće inteligencije. Za njeno stvaranje 1983. godine zaslužni su John Laird, Allen Newell i Paul Rosenbloom. Soar pohranjuje svoja znanja u obliku produkcijskih pravila raspoređeni u smislu operatora koji djeluju na prostoru problema, odnosno na skupu stanja koja predstavljaju neki zadatak. U arhitekturu je također ugrađen jedan mehanizam analitičkog učenja nazvan komadanje (*chunking*) kao varijanta spomenute EBL metode. Svako produkcijsko pravilo koje nastane komadanjem slijedi deduktivno iz prethodnog pravila u sustavu što znači da ako prethodno znanje nije točno, nova pravila također mogu biti netočna. Za slučajeve pojave kontradiktornih pravila nastalih na temelju netočnog ili interno nekonzistentnog znanja Soar posjeduje mehanizam za korekciju koji uspoređuje ispravnost tog pravila s drugim pravilom koje se primjenjuje u sličnoj situaciji. Hoće li ovakav mehanizam u konačnici biti implementiran ovisi o samom obliku početnog znanja, ali i o redoslijedu u kojem su primjeri predstavljeni. Laički rečeno, ovakav mehanizam omogućuje provjeru neispravnih pravila učenjem novih koja se bolje uklapaju u dostupne podatke.⁴⁵ U posljednjih nekoliko godina arhitekturi su dodane mnoge ekstenzije poput učenja podrškom koje je namijenjeno za

⁴⁴ Ibid.

⁴⁵ Mitchell, T. M.; Thrun, S. B. Learning Analytically and Inductively, u: *Mind Matters: A Tribute to Alan Newell*, uredili Steier, D. M. i Mitchell, T. M. Erlbaum, 1996. Str. 85-111.
URL: https://www.ri.cmu.edu/pub_files/pub1/mitchell_tom_1996_1/mitchell_tom_1996_1.pdf (20.8.2016.)

prilagođavanje preferiranih vrijednosti operatora, epizodno učenje kako bi se zadržala povijest evolucije sustava, semantičko učenje za opisivanje apstraktnog deklarativnog znanja itd. Sustav je relativno uspješno demonstrirao niz visoko razinskih kognitivnih funkcija kao što su obrada velikih i složenih setova pravila u planiranju, rješavanju problema i razumijevanju prirodnog jezika u realnom vremenu.⁴⁶

4.1.2.3. EPIC

EPIC (*Executive Process Interactive Control*) je također jedan primjer kognitivne arhitekture namijenjene za izgradnju računalnog modela koji bi sadržavao određene aspekte ljudske inteligencije. Arhitektura je razvijena pod vodstvom Davida Kieras i Davida Meyera, a njezin cilj je dvostruki. S jedne strane, ona pokušava obuhvatiti ljudske kognitivne i motoričke sposobnosti kroz nekoliko međusobno povezanih procesora koji rade paralelno, a s druge strane također nastoji izgraditi modele ljudsko-računalne interakcije za praktične svrhe. Kontrola sustava vrši se produkcijskim pravilima za kognitivni procesor te skupom perceptivnih (vizualni, slušni i taktilni procesor) i motornih procesora koji rade na simbolički kodiranim značajkama. Osnovno područje kojim se sustav bavi su situacije u kojima je potrebno izvršiti nekoliko zadataka istovremeno (npr. ispijanje kave prilikom vožnje auta). Korištenjem EPIC-a moguće je konstruirati modele koji predstavljaju generalne procedure potrebne kako bi se izvršili takvi kompleksni multimodalni zadaci u obliku skupa proceduralnih pravila. Njegova specifičnost leži u tome što za izvršavanje zadataka arhitektura specifično zahtijeva i produkcijski dio programa kao i relevantne perceptualno-motoričke parametre. Bitno je naglasiti kako se ne radi o sustavu za učenje te sam program ne posjeduje nikakvu vrstu takvih mehanizama. Štoviše, njegova svrha je detaljna reprezentacija perceptualnih, motoričkih i kognitivnih ograničenja ljudske sposobnosti. No isto tako u jednom eksperimentu povezan je sa Soar arhitekturom za potrebe rješavanja problema, planiranja i učenja, a takva EPIC-SOAR kombinacija čak je uspješno primijenjena na simulaciji kontrole zračnog prometa.⁴⁷

⁴⁶ Duch, W.; Oentaryo, R. J.; Pasquier, M. Cognitive Architectures: Where do we go from here?, u: *Frontiers in Artificial Intelligence and Applications*, vol. 171, uredili Wang, P; Goertzel, B. i Franklin, S. IOS Press, ožujak 2008. Str. 122-136. URL: <http://www.fizyka.umk.pl/publications/kmk/08-AGI.pdf> (20.8.2016.)

⁴⁷ Kieras, D. E.; Meyer, D. E. An Overview of the EPIC Architecture for Cognition and Performance With Application to Human-Computer Interaction, u: *Human-Computer Interaction*, vol. 12, br. 4, prosinac 1997. Str. 391-438. URL: <http://umich.edu/~bcalab/documents/KierasMeyer1997.pdf> (20.8.2016.)

4.2. Podsimbolički pristup

Alternativa prethodnom pristupu okvirno je poznata kao podsimbolička paradigma. U ovoj paradigmi postoji srednja razina strukture koja se nalazi između neuronske i simboličke razine. Ova podsimbolička razina je pokušaj formaliziranja, na nekoj razini apstrakcije, vrste procesiranja koja se odvija u živčanom sustavu. Osnovna karakteristika ovog pristupa je očekivanje kako će apstraktna manipulacija simbola, uz svaki drugi aspekt inteligencije, proizaći iz nižerazinske podsimboličke dinamike. Većina ovakvih sustava inspirirana je konekcionistačkim idejama, a njezine korijene moguće je pronaći u teoriji kibernetike Norberta Wienera iz 1948. ili konceptu perceptrona Franka Rosenblatta iz 50- i 60-ih godina. Iako je ovo područje nekoliko puta padalo u zaborav, prava revolucija na području konektivizma nastupila je početkom 1980-ih s formulacijom ideje paralelnog distribuiranog procesiranja (*parallel distributed processing* - *PDP*). U osnovi, radilo se o konceptu neuronskih mreža koji je naglašavao paralelnu prirodu neuronske obrade informacija, ali i distribuiranu prirodu neuronskih reprezentacija. Umjetne neuronske mreže bazirane na ovom principu su biološki inspirirani računalni modeli te se sastoje od elemenata za obradu (predstavljaju neurone) i veza (predstavljaju sinapse) između njih kojima su pridruženi koeficijenti, odnosno težine. Većina mreža prima vektore ili druge jednostavne strukture podataka kao ulaz i proizvodi opet relativno jednostavne strukture podataka kao izlaz. Podaci se pohranjuju implicitno kroz parametre mreže, najčešće kroz težinu veza između neurona, a obrada podataka se postiže širenjem aktivnosti kroz mrežu (*propagation*).⁴⁸ Sami sustavi mogu biti podijeljeni na nekoliko slojeva (ulazni, izlazni i skriveni) koji mogu sadržavati različite količine neurona.

Do danas je razvijen veliki broj mreža koje se mogu razlikovati po broju neurona, vrstama veza, slojevima, korištenim algoritmima itd. Budući da bi za potrebe ovoga rada bilo nemoguće prikazati sve tipove, za primjer može poslužiti koncept klasične neuronske mreže (*classical neural network* - *CNN*). Iako se radi o najjednostavnijem primjeru, ovakav sustav koristi fundamentalne ideje konekcionistačkog područja koje su implementirane u većini današnjih mreža. Glavni aspekti klasičnog modela su sljedeći:

⁴⁸ Kurfess, F. J. Integrating symbol-oriented and subsymbolic reasoning methods into hybrid systems, u: *From Synapses to Rules - Discovering Symbolic Rules from Neural Process Data*, uredili Apolloni, B. i Kurfess, F. J. Kluwer Academic Publishers, 2002. Str. 275–292.

URL: http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1016&context=csse_fac (20.8.2016.)

1. Mreža se sastoji od konstantnog broja neurona te veza između njih. Svaki neuron ima aktivacijsku vrijednost, a svaka veza vrijednost težine. Sami neuroni mogu biti ulazni, izlazni ili skriveni što ovisi o njihovoj vezi s okolinom. Stanje mreže u bilo kojem trenutku može se opisati vektorom duljine N koji onda može biti podijeljen u ulazni, izlazni i skriveni vektor. Sve vrijednosti težine u mreži u bilo kojem trenutku mogu se opisati s $N \times N$ matricom.
2. U svakom koraku promjene stanja neuron dobiva svoj ulaz od ostalih neurona ili iz vanjskog okruženja. Njegova aktivacijska funkcija zatim pretvara ovaj ulaz u svoju ažuriranu aktivacijsku vrijednost koja je ujedno i njegova izlazna vrijednost te ju šalje drugim neuronima ili vanjskom okruženju. Kao rezultat tog procesa vektor stanja mijenja svoju vrijednost, a ovaj postupak se ponavlja sve dok vektor stanja ne poprimi stabilnu vrijednost.
3. Kroz fazu treniranja mreže težine veza usklađuju se korištenjem nekog algoritma za učenje. Primjeri tijekom treninga moraju se kontinuirano predstavljati sustavu sve dok matrica težina ne poprimi stabilnu vrijednost.
4. Kao posljedica treninga mreža postaje funkcija koja mapira ulazne vektore na izlazne vektore ili vektore stanja. Točnije, polazeći od početnog stanja širenje aktivacijskog procesa će dovesti sustav do stabilnog izlaznog vektora (kada se mreža koristi kao funkcija) ili stabilnog vektora stanja (kada se mreža koristi kao asocijativna memorija).⁴⁹

Kao što je spomenuto postoje različiti tipovi mreža koji se međusobno razlikuju po strukturi, aktivacijskoj funkciji, algoritmima za učenje itd. No u odnosu na njihovu arhitekturu postoje dva osnovna tipa. Prvi čine mreže bez povratne veze (*feedforward networks*). Njihova osnovna karakteristika je ta što ne posjeduju veze koje idu natrag od izlaznih prema ulaznim neuronima, a mreža ne zadržava memoriju svojih prethodnih izlaznih vrijednosti kao ni aktivacijska stanja svojih neurona. U ovakvim mrežama informacije se kreću samo prema naprijed što znači da ne postoje veze između neurona u istome sloju kao ni petlje ili ciklusi u procesu izračunavanja.⁵⁰ Ovakav dizajn je poprilično jednostavan i donekle ograničen, ali također rezultira s predvidljivim vremenom izračunavanja neke funkcije. Višeslojne mreže

⁴⁹ Wang, P. Artificial General Intelligence and Classical Neural Network, u: *Proceedings of the IEEE International Conference of Granular Computing*. IEEE Computational Intelligence Society, Atlanta, Georgia, svibanj 2006. Str. 130-135. URL: <http://cis-linux1.temple.edu/~pwang/Publication/AGI-CNN.pdf> (20.8.2016.)

⁵⁰ Russell, S. J.; Norvig, P. Artificial Intelligence: A Modern Approach. Prentice-Hall, Englewood Cliffs, 1995. URL: <http://stpk.cs.rtu.lv/sites/all/files/stpk/materiali/MI/Artificial%20Intelligence%20A%20Modern%20Approach.pdf> (20.8.2016.)

bez povratne veze pokazale su se iznimno efikasnim u prepoznavanju uzoraka. Drugi tip čine povratne ili rekurzivne mreže (*reccurent networks*) gdje poveznice između neurona mogu oblikovati bilo koju proizvoljnu topologiju. Takva mreža čuva memoriju svojih ranijih stanja, a sljedeće stanje ne ovisi samo o ulaznim signalima već i o prethodnim stanjima mreže. Zbog svoje složene unutarnje strukture povratne mreže su sposobne za sofisticiranije unutarnje reprezentacije, ali zato zahtijevaju kompliciranije metode učenja.⁵¹

4.2.1. Prikaz znanja

U usporedbi sa simboličkim pristupom, relativno je teško izraditi sofisticirane metode prikaza znanja u konekcionističkim modelima. Općenito gledajući, postoje dva osnovna pristupa. U prvome svaki neuron posjeduje određeno individualno značenje u svom okruženju. Na primjer, ulazni neuron može odgovarati jednom atributu objekta koji se klasificira, a izlazni neuron može odgovarati klasi u koju se objekt klasificira. To znači da konceptualni entiteti odgovaraju pojedinim neuronima, a svaki neuron je povezan s prikazom samo jednog entiteta. Odnosi ovdje mogu biti direktno izraženi vezom između pojedinih neurona. U ovakvoj situaciji sustav u osnovi postaje sredstvo za izraz semantičke mreže ili neke slične reprezentacije na bazi grafova. Ovo je poznato kao lokalna reprezentacija te je relativno eksplicitna i lako razumljiva, a može biti generirana iz drugih metoda prikaza pomoću procesa transformacije ili kompilacije.⁵²

Drugi pristup spada pod distribuiranu reprezentaciju gdje neuron nema individualno značenje već su ulazno-izlazni vektori minimalne značenjske jedinice u sustavu, a svaki neuron pridonosi predstavljanju nekoliko entiteta. Trenutačno postoji nekoliko načina za prikazivanje logike i pravila u distribuiranim konekcionističkim modelima. Jedan primjer bio bi reprezentacija pravila kroz korištenje direktnih poveznica između neurona koji predstavljaju uvjete s neuronima koji predstavljaju zaključke. Proces zaključivanja u ovakvom modelu je sveden na širenje aktivnosti.⁵³ Karakteristike distribuiranog pristupa najlakše je opisati na primjeru klasične neuronske mreže (*CNN*) gdje se kao najveća prednost ističe sposobnost asocijativne memorije. Naime, nakon što se neki uzorak pružen tijekom treninga pohrani u matricu težina, moguće ga je adresirati sadržajem. To znači da će pojava

⁵¹ Kurfess, F. J. Integrating symbol-oriented and subsymbolic reasoning methods into hybrid systems, u: *From Synapses to Rules - Discovering Symbolic Rules from Neural Processsed Data*, uredili Apolloni, B. i Kurfess, F. J. Kluwer Academic Publishers, 2002. Str. 275–292.

URL: http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1016&context=csse_fac (20.8.2016.)

⁵² Ibid.

⁵³ Ibid.

novog podražaja natjerati mrežu da poprими aktivacijski uzorak primjera koji je najbližiji novom podražaju. Na primjer, ako se trening sastoji od niza fotografija, a novi podražaj je mali komad jedne od njih, mreža će reproducirati aktivaciju fotografije iz koje je komad uzet. Izvorne fotografije ovdje nisu pohranjene odvojeno u mreži već svaka težina predstavlja djelomično kodiranje svih fotografija. Za ovakve potrebe najčešće se koriste poznate Hopfield mreže. Druga velika prednost je mogućnost toleriranja greški. Ako je dio mreže uništen, ona ne gubi svoju memoriju zato što se pohranjeni uzorci uglavnom mogu dohvatiti, doduše s manjom točnošću. Ovo je poznato kao elegantna degradacija (*graceful degradation*).⁵⁴ Za usporedbu, kod sustava s lokalnim reprezentacijama informacije u oštećenim dijelovima će biti potpuno izgubljene. Očita pozitivna strana je ta što se neuronske mreže mogu nositi s nepotpunim, nekonzistentnim ili aproksimativnim informacijama.

No s druge strane postoje određeni nedostaci sa stajališta zahtjeva umjetne opće inteligencije. Među njima se ističe pitanje strukturiranog znanja. Iako se u teoriji komad znanja uvijek može kodirati vektorom, nije baš jednostavno predstaviti njegovu strukturu kada je sastavljena od manjih komada. Jedno od mogućih rješenja je recimo kodiranje skripti kroz podjelu ulaznih jedinica u segmente od kojih svaki kodira aspekt skripti u distribuiranom obliku. Drugi problem tiče se selektivne obrade. Budući da su svi dijelovi znanja pomiješani u matrici težina, nije lako obraditi neke od njih selektivno bez dodirivanja drugih. Za općeniti sustav takva selektivna obrada je neophodna zato što će u različitim situacijama samo relevantno znanje biti upotrijebljeno. I zadnji problem odnosi se na mogućnost objašnjavanja. Naime, dobro je poznata činjenica kako je iznimno teško dobiti smislena objašnjenja unutarnjih procesa neuronskih mreža. Iako je donekle moguće objasniti zašto mreža stvara određeni rezultat u smislu proračuna vektora stanja, to nije ono što ljudi mogu intuitivno shvatiti. Suvislo objašnjenje obično podrazumijeva opis procesa na sažet način s poznatim pojmovima što se čini nespojivim sa samom idejom distribuirane reprezentacije.⁵⁵

4.2.2. Učenje

Za razliku od simboličke umjetne inteligencije koja je od početka bila usmjerena na prikaz znanja, sam temelj konekcionističkog modela oduvijek je bilo učenje. Učenje ovdje uglavnom podrazumijeva podešavanje težina ili drugih parametara kako bi se složeni

⁵⁴ Wang, P. Artificial General Intelligence and Classical Neural Network, u: *Proceedings of the IEEE International Conference of Granular Computing*. IEEE Computational Intelligence Society, Atlanta, Georgia, svibanj 2006. Str. 130-135. URL: <http://cis-linux1.temple.edu/~pwang/Publication/AGI-CNN.pdf> (20.8.2016.)

⁵⁵ Ibid.

proračuni mogli ostvariti širenjem aktivacija kroz te težine. Podešavanje težina postiže se primjenom nekog algoritma za učenje. Među najstarije ovakve algoritme spada Hebbovo učenje (*Hebbian learning*) nazvano po Donaldu Hebbu. Ideja ovog generičkog principa u velikoj mjeri temelji se na dinamici bioloških sustava u smislu da sinapse koje povezuju dva neurona povećavaju svoju snagu ako se oba neurona kontinuirano istovremeno aktiviraju na ulazne podražaje.⁵⁶ U skladu s time, koristeći ovaj pristup unutar neuronske mreže težina veza se povećava ako postoji visoki stupanj korelacije između dva povezana neurona. Možda i najpopularnija paradigma koja se koristi u neuronskim mrežama je algoritam s povratnim postupkom (*backpropagation algorithm*). On za svaki par ulaznog uzorka i željenog odgovora izračunava trenutnu reakciju mreže prema konfiguraciji težina te uspoređuje rezultat s željenim odgovorom, a razlika između ove dvije vrijednosti se koristi za podešavanje težina.⁵⁷ Proces izračuna pogreške za svaki neuron vraća se natrag u mrežu u novi korak te se ponavlja sve dok pojedinačna ili ukupna pogreška nije niža od postavljene granične vrijednosti čime je učenje mreže službeno završeno.

U smislu općenitog načela učenja, konekcionistički algoritmi su osmišljeni za tri vrste pristupa:

1. Nadgledano učenje (*supervised learning*) sadrži nekakvu vrstu poticaja za vrijeme treniranja mreže. Svaki primjer koji je pružen tijekom treninga je par koji se sastoji od ulaznog objekta (vektora) te željene izlazne vrijednosti. Ovu konačnu izlaznu vrijednost ocjenjuje nadzornik ili trener, uglavnom u obliku algoritma, kako bi pomoću ispravljanja pogrešaka mijenjao težinu veza te samim time mijenjao način na koji mreža reagira. Primjer bi bio algoritam s povratnim postupkom.
2. Nenadgledano učenje (*unsupervised learning*) je oblik strojnog učenja čiji je zadatak izračunavanje neke funkcije na temelju neobilježenih podataka. Budući da ovdje nije prisutan nikakav algoritam koji nadgledava i usmjerava proces, ne postoji ni eksplicitna željena izlazna vrijednost ili evaluacija potencijalnog rješenja. Dobar primjer bilo bi Hebbovo učenje.
3. Učenje podrškom (*reinforcement learning*) je u određenoj mjeri kombinacija prethodna dva pristupa. Poput nenadgledanog učenja ovdje ne postoji nadzornik koji

⁵⁶ Hebbian theory. Wikipedia. URL: https://en.wikipedia.org/wiki/Hebbian_theory (20.8.2016.)

⁵⁷ Kurfess, F. J. Integrating symbol-oriented and subsymbolic reasoning methods into hybrid systems, u: *From Synapses to Rules - Discovering Symbolic Rules from Neural Processsed Data*, uredili Apolloni, B. i Kurfess, F. J. Kluwer Academic Publishers, 2002. Str. 275–292.
URL: http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1016&context=csse_fac (20.8.2016.)

ispravlja rezultate i usmjerava rad mreže, ali zato postoji određena vrsta poticaja za vrijeme treniranja. U ovome pristupu poticaji su implementirani u obliku pozitivnih i negativnih povratnih informacija koje mreža prima iz okoline. Na temelju tih informacija i indeksa kojim se mjeri kvaliteta učenja, mreža usvaja koncepte metodom pokušaja i pogrešaka.

Učenje podrškom je jedna od najpopularnijih paradigmi kod neuronskih mreža. Za razliku od nadgledanog učenja, korištenjem ovog pristupa mreže bez nadzornika mogu naučiti nove koncepte koji nisu pokriveni primjerima za vrijeme treninga. No i ovaj princip učenja posjeduje određene nedostatke. Osnovni problem predstavlja veličina prostora stanja za kompleksnije probleme. Laički rečeno, u većini situacija količina potencijalnih akcija koje sustav može poduzeti je zaista ogromna što može znatno usporiti proces učenja. Iz tog razloga ova metoda se uglavnom koristi s nekom vrstom optimizacijskog algoritma, kao što je *Q-learning*. Ovaj algoritam koristi Q vrijednost kao evaluaciju kvalitete potencijalne akcije u nekom stanju, odnosno indikaciju koliko je ta akcija poželjna. Princip je načelno jednostavan. Naime, sustav održava bazu podataka u koju zapisuje parove stanja i mogućih akcija, a svakome paru je pridružena Q vrijednost. Ako za poduzetu akciju sustav primi pozitivnu povratnu informaciju, njezina Q vrijednost se povećava. U slučaju negativne informacije vrijednost se smanjuje.⁵⁸

Učenje u konektivističkom području također je uspješno primijenjeno s određenim oblicima simboličkog znanja. Na primjer, moguće je iskoristiti prethodno opisani algoritam s povratnim postupkom za učenje strukture stabla kroz kontinuiranu primjenu algoritma na različitim mjestima grananja.⁵⁹ Jako dobar primjer je mogućnost ekstrakcije pravila iz neuronskih mreža. Prema jednom mogućem postupku za ovakav proces koristi se model s dvije razine. Donja razina implementira opisano Q učenje u mreži s povratnim postupkom dok gornja razina sadrži algoritam baziran na učenju s podrškom. Učenje pravila odvija se na svakom koraku koje je povezano s podacima x , y , r i a , gdje je x stanje prije akcije a , y je novo stanje, a r predstavlja povratnu informaciju. Ako je za neku poduzetu akciju na donjoj razini na temelju povratnih informacija procijenjeno da je uspješna, onda algoritam izdvaja pravilo koje odgovara toj odluci te ga dodaje popisu postojećih pravila. U kasnijim

⁵⁸ Sun, R. Supplementing Neural Reinforcement Learning with Symbolic Methods, u: *Hybrid Neural Systems*. Springer-Verlag, London, UK, prosinac 1998. Str. 333-347. URL: www.cogsci.rpi.edu/~rsun/sun.hybrid99.ps (20.8.2016.)

⁵⁹ Sun, R. Artificial Intelligence: Connectionist and Symbolic Approaches, u: *International Encyclopedia of the Social and Behavioral Sciences*, uredili Smelser, N. J. i Baltes, P. B. Pergamon/Elsevier, Oxford, 2001. Str. 783-789. URL: <http://www.cogsci.rpi.edu/~rsun/sun.encyc01.pdf> (20.8.2016.)

interakcijama s okolinom algoritam provjerava izvađeno pravilo u smislu ishoda njegove primjene. Ako rezultat nije uspješan onda se pravilo prilagođava kako bi bilo više specifično i ekskluzivno za pojedinačni slučaj. Ovakav postupak naziva se smanjivanje (*shrinking*). U suprotnom pravilo se može širiti (*expansion*), odnosno ako je rezultat njegove primjene uspješan može ga se generalizirati kako bi bio više univerzalan. Pravila su u standardnom obliku uvjeta i akcija.⁶⁰

Što se tiče karakteristika potrebnih za opće inteligentne sustave, proces učenja u neuronskim mrežama posjeduje sljedeće prednosti:

- Algoritmi opće namjene: algoritmi učenja u klasičnoj neuronskoj mreži dizajnirani su na način koji je neovisan o specifičnoj domeni. Ovakvi sustavi ne zahtijevaju ugrađeno znanje o nekom području za razliku od simboličkih ekspertnih sustava
- Generalizacija i sličnost: klasična mreža automatski generalizira naučene uzorke kako bi pokrila slične primjere. Štoviše, ona može raditi s uzorcima koji joj prije nisu bili dostupni dok god posjeduju nekakvu sličnost s naučenim primjerima
- Toleriranje šumova: naučene pravilnosti u mreži nisu narušene pojavom kontradiktornih primjera već se ovakvi primjeri označavaju kao šum ili buka u podacima te ih se ignorira. Isto tako, ako neki uzorak sadrži određene vrijednosti koje se ne uklapaju, mreža ih promatra kao šum te umjesto njih koristi standardne vrijednosti. S druge strane, ovakav učinak jako je teško postići korištenjem binarne logike⁶¹

Naravno i konekcionistački pristup pokazuje određene nedostatke. Sa stajališta UOI sustava postoji nekoliko problema. Prvi je vezan uz inkrementalno učenje. Naime, klasična mreža pretpostavlja da su svi uzorci dostupni na početku faze učenja. Kada ovi uzorci postanu dostupni s vremena na vrijeme, kao što je slučaj s cjeloživotnim učenjem, mrežu je potrebno ponovno istrenirati sa svim novim i starim uzorcima što naravno predstavlja nerealne vremenske zahtjeve. Nažalost, rješenje nije inkrementalno prilagođavanje istrenirane mreže zato što takav postupak često dovodi do uništenja stare memorije. Ovaj problem poznat je pod nazivom katastrofalno zaboravljanje (*catastrophic forgetting*). Drugi nedostatak odnosi se na

⁶⁰ Sun, R. Supplementing Neural Reinforcement Learning with Symbolic Methods, u: *Hybrid Neural Systems*. Springer-Verlag, London, UK, prosinac 1998. Str. 333-347. URL: www.cogsci.rpi.edu/~rsun/sun.hybrid99.ps (20.8.2016.)

⁶¹ Wang, P. Artificial General Intelligence and Classical Neural Network, u: *Proceedings of the IEEE International Conference of Granular Computing*. IEEE Computational Intelligence Society, Atlanta, Georgia, svibanj 2006. Str. 130-135. URL: <http://cis-linux1.temple.edu/~pwang/Publication/AGI-CNN.pdf> (20.8.2016.)

nemogućnost raznolikog učenja. Mreže usvajaju koncepte isključivo na temelju konkretnih primjera dok proces učenja kod ljudi postoji u različitim formama. Na primjer, novo znanje moguće je usvojiti u obliku općih izjava, a novi zaključci mogu nastati procesima koji su kompliciraniji od obične indukcije ili analogije. Treći veliki problem je pitanje iznimki. Iako je sposobnost mreži da ignoriraju iznimke u određenoj mjeri prednost, ovakav pristup nije najpoželjniji kod UOI sustava. Oni bi naime trebali pamtit i opće tendencije među uzorcima i njihove iznimke, a onda odlučivati koje je potrebno primjeniti u novim situacijama. I na kraju, jedna od općenitijih zamjerki je sposobnost učenja u stvarnom vremenu. Dobro je poznata činjenica kako proces učenja u neuronskim mrežama može biti veoma dugotrajan.⁶²

Kao nekakav konačan zaključak, konekcionistički pristup omogućuje učenje veoma kompliciranih funkcija koje je nemoguće postići drugim UI pristupima. No isto tako veoma je ograničen po pitanju postizanja višestrukih ciljeva. Trenutačno nije moguće izvršavati nekoliko posve različitih funkcija korištenjem jedne neuronske mreže. Također, izvršavanje funkcije koja stalno postiže isti rezultat možda je efikasno, ali nije najbolji primjer prave inteligencije.

4.2.2.1. IBCA

IBCA (*Integrated Biologically-based Cognitive Architecture*) je velika kognitivna arhitektura koja utjelovljuje automatsku i distribuiranu prirodu obrade informacija u mozgu. Naglašene su tri specifične regije mozga: stražnji korteks (*posterior cortex - PC*), frontalni korteks (*frontal cortex - FC*) i hipokampus (*hippocampus - HC*). PC modul koristi dualnu lokalno-distribuiranu reprezentaciju koja se fokusira na obradu senzo-motoričkih informacija. FC modul koristi lokalnu reprezentaciju u povratnoj mreži u kojoj su jedinice radne memorije izolirane jedna od druge te doprinose kombinatorno u smislu da zasebne aktivne jedinice predstavljaju različite značajke. HC modul koristi potpuno globalnu distribuiranu organizaciju neurona gdje sve jedinice doprinose interaktivno određenoj reprezentaciji. Ovakva struktura omogućuje brzo povezivanje svih aktivacijskih uzoraka u PC i FC modulu što predstavlja ulogu epizodnog pamćenja. Algoritmi za učenje uključuju LEABRA algoritam za usvajanje vještina kroz proces ispravljanja pogrešaka te Hebbovo učenje. PC i FC moduli koriste sporo integrativno učenje koje spaja mnoga pojedinačna iskustva kako bi obuhvatili temeljne uzorke u svojoj okolini te kako bi podržali senzorno-motoričke aktivnosti dok HC modul implementira brzo učenje koje zadržava i razlikuje pojedinačna iskustva. Trenutačno se

⁶² Ibid.

smatra kako ovakva konfiguracija i suradnja između modula odražava paradigme učenja prisutne u ljudskom mozgu. Komplementarnu tehniku učenja predstavlja konsolidacija znanja, odnosno prijenos kratkoročnog u dugoročno pamćenje. Sadašnja arhitektura je ograničena na učenje parametara težine, ali ne i lokalne strukture mreže. Iako se ovakvim tipom arhitekture može objasniti ljudsko ponašanje u određenim segmentima, još uvijek nije demonstrirano kako bi se mogla koristiti za zadatke koji zahtijevaju rasuđivanje.⁶³

4.2.2.2. CORCOTRONICS

Corcotronics je podsimbolička arhitektura koja modelira biološke funkcije moždane kore i talamusa. Njezina organizacija memorije sastoji se od jedinica koje se nazivaju leksikoni. Svaki leksikon sadrži jedan lokalni dio kore, lokalni dio talamusa te njihove međusobne poveznice. Također, svaki leksikon može poprimiti različita stanja koja se nazivaju simboli, a predstavljeni su sa skupom različitih neurona. U skladu s time, znanje u ovoj arhitekturi nalazi se u obliku paralelnih jednosmjernih veza između neurona koji predstavljaju jedan simbol u leksikonu s onima koji opisuju simbol u drugom leksikonu. Svaka od tih veza predstavlja neku jedinicu znanja, a skup svih veza čini bazu znanja. Za proces učenja, odnosno aktivaciju neurona koristi se takozvana kompetitivna metoda u kojoj se neuroni međusobno natječu kako bi postali aktivni. Ovdje se radi o nenadgledanoj vrsti učenja u kojoj se u svakom trenutku dopušta samo jednom neuronu da bude aktivan. Iako ova arhitektura uspješno modelira procese mašte i kreativnosti još uvijek nije osposobljena za kompleksne procese zaključivanja.⁶⁴

4.2.2.3. DeSTIN

DeSTIN je primjer kompleksne neuronske mreže namijenjene za prepoznavanje vizualnih uzoraka. Sama mreža je hijerarhijske strukture te se sastoji od nekoliko različitih slojeva. Čvorovi na najnižim razinama kao ulaz primaju sirove podatke u obliku piksela slike. Drugi sloj, ali i svi oni iznad njega, kao ulaz primaju stanja čvorova na odgovarajućim nižim slojevima. Također, čvorovi na višim razinama odgovaraju većim regijama prostora i vremena, a svaki od njih izvodi predviđanja u vezi događaja u prostorno-vremenskom području na koje se odnosi. Kombiniranjem jednosmjernih i povratnih informacija s predviđanjima unutar čvorova stvara se kompleksni nelinearni dinamički sustav čije se stanje

⁶³ Duch, W.; Oentaryo, R. J.; Pasquier, M. Cognitive Architectures: Where do we go from here?, u: *Frontiers in Artificial Intelligence and Applications*, vol. 171, uredili Wang, P; Goertzel, B. i Franklin, S. IOS Press, ožujak 2008. Str. 122-136. URL: <http://www.fizyka.umk.pl/publications/kmk/08-AGI.pdf> (20.8.2016.)

⁶⁴ Ibid.

samostalno organizira na način da odražava stanje u svijetu koji se percipira. Arhitektura je sposobna za veoma učinkovitu nenadgledanu klasifikaciju različitih vizualnih ulaznih podataka. Trenutačno je u stanju kreirati strukture koje odgovaraju različitim kategorijama objekata, kao na primjer krevet, lampa, kućni ljubimac itd. No do danas DeSTIN je isključivo korišten samo na obradi vizualnih i zvučnih informacija.⁶⁵

4.3. Hibridni pristup

Posljednju bitnu paradigmu u izradi UOI sustava predstavlja takozvani hibridni ili integrativni pristup. Ovo podrazumijeva uzimanje nekih ili svih elemenata od gore navedenih pristupa te stvaranje kombiniranih sinergističkih sustava. Naravno, ovakav koncept ima smisla ako se prihvati pretpostavka da drugačiji UI pristupi zahvaćaju neki jedinstveni aspekt ljudske inteligencije. Na primjer, simboličke arhitekture su bolje u ostvarivanju kognitivnih funkcija visoke razine, kao što su planiranje i zaključivanje, te posjeduju sposobnosti veoma slične ljudskoj ekspertizi. Međutim, njihov najveći problem su formulacija simboličkih entiteta iz niskorazinskih informacija kao i funkcioniranje s velikim količinama nepotpunih i parcijalnih informacija. S druge strane, konekcionistički sustavi su bolje prilagođeni za obuhvaćanje specifičnog konteksta ljudskih sposobnosti kao i za mogućnost funkcioniranja s velikom količinom niskorazinskih informacija, dok im je glavna mana nemogućnost ostvarivanja kognitivnih funkcija više razine.⁶⁶ Samu integraciju komponenti moguće je obaviti na nekoliko različitih načina, no osnovni problem u izradi ovakvih sustava je konfiguracija različitih modula koji predstavljaju različite UI paradigme.

Što se tiče prikaza znanja kod hibridnog pristupa, ono može biti u dva najosnovnija oblika. Prvi je lokalno-distribuirani što znači da se sustav sastoji od kombinacije lokalnih modula gdje je svaki koncept definiran s jednim čvorom i distribuiranih modula gdje je svaki koncept predstavljen pomoću skupa različitih čvorova. Drugi pristup je simboličko-konekcionistički. On uključuje spajanje različitih simboličkih metoda prikaza, poput pravila ili grafova, te konekcionističkih metoda lokalnog ili distribuiranog tipa. Za potrebe učenja ovakvih sustava također postoji nekoliko mogućnosti. Jedna od njih je recimo prethodno opisani postupak izvlačenja simboličkog znanja u obliku pravila iz neuronske mreže koja

⁶⁵ Arel, I.; Rose, D.; Coop, R. DeSTIN: A Scalable Deep Learning Architecture with Application to High-Dimensional Robust Pattern Recognition. AAAI Fall Symposium Series, siječanj 2009.

URL: <http://web.eecs.utk.edu/~itamar/Papers/BICA2009.pdf> (20.8.2016.)

⁶⁶ Duch, W.; Oentaryo, R. J.; Pasquier, M. Cognitive Architectures: Where do we go from here?, u: *Frontiers in Artificial Intelligence and Applications*, vol. 171, uredili Wang, P; Goertzel, B. i Franklin, S. IOS Press, ožujak 2008. Str. 122-136. URL: <http://www.fizyka.umk.pl/publications/kmk/08-AGI.pdf> (20.8.2016.)

koristi učenje s podrškom. Ta pravila se onda mogu dodatno iskoristiti za ubrzavanje procesa učenja kao i za poboljšanje prijenosa znanja između različitih modula. U hibridnim sustavima moguće je koristiti cijeli niz različitih metoda za potrebe učenja kao što su razni algoritmi za pretraživanje, evolucijski algoritmi, algoritam postupnog opadanja (*gradient descent*) ili pak razne heurističke metode poput stabla odlučivanja ili indukcije pravila.⁶⁷ Prema samom smjeru učenja hibridne arhitekture može se kategorizirati u dvije glavne skupine, učenje odozgo (*top-down*) i učenje odozdo (*bottom-up*). Prvi pristup uključuje prijelaz znanja s eksplicitne konceptualne razine na implicitnu. Ovo je moguće postići prethodnim kodiranjem skupa ekspertnih pravila na višoj razini simboličkog ili lokalnog modula te omogućavanjem nižoj razini koja se sastoji od distribuirane neuronske mreže da uči promatranjem aktivnosti tih pravila. S druge strane, učenje odozdo može se postići vađenjem ili prevođenjem implicitnog znanja kodiranog u nižerazinskim modulima u skup konceptualnih pravila. Primjer ovoga je opet prethodno opisani postupak ekstrakcije pravila iz neuronskih mreža.⁶⁸

Među bitnije aspekte izrade hibridnih sustava svakako spada i način konfiguracije različitih modula, odnosno stupanj integracije komponentni. Prema ovome kriteriju moguće je definirati pet različitih tipova ili pristupa izradi hibridnih sustava:

1. Samostalna arhitektura bazira se na integraciji nezavisnih komponentni koje koriste različite metode. Uporaba već postojećih komponenti, bilo softverskih ili hardverskih, nudi prednost jednostavne implementacije, osobito u najekstremnijim slučajevima gdje ne postoji direktna interakcija između komponenti. Druge prednosti uključuju redundanciju u slučaju kvara te mogućnost validacije u smislu da se jedna komponenta koristi za provjeru rezultata druge komponente. Glavne mane ovakvog pristupa su nedostatak prijenosa podataka između komponenti, potreba višestrukog održavanja sustava, pogotovo ako je isto znanje prikazano u različitim komponentama, te mogući manjak dosljednosti sustava.⁶⁹

⁶⁷ Sun, R. Artificial Intelligence: Connectionist and Symbolic Approaches, u: *International Encyclopedia of the Social and Behavioral Sciences*, uredili Smelser, N. J. i Baltes, P. B. Pergamon/Elsevier, Oxford, 2001. Str. 783-789. URL: <http://www.cogsci.rpi.edu/~rsun/sun.encyc01.pdf> (20.8.2016.)

⁶⁸ Duch, W.; Oentaryo, R. J.; Pasquier, M. Cognitive Architectures: Where do we go from here?, u: *Frontiers in Artificial Intelligence and Applications*, vol. 171, uredili Wang, P; Goertzel, B. i Franklin, S. IOS Press, ožujak 2008. Str. 122-136. URL: <http://www.fizyka.umk.pl/publications/kmk/08-AGI.pdf> (20.8.2016.)

⁶⁹ Kurfess, F. J. Integrating symbol-oriented and subsymbolic reasoning methods into hybrid systems, u: *From Synapses to Rules - Discovering Symbolic Rules from Neural Process Data*, uredili Apolloni, B. i Kurfess, F. J. Kluwer Academic Publishers, 2002. Str. 275–292. URL: http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1016&context=csse_fac (20.8.2016.)

2. Transformacijski pristup koristi se pretvorbom između konvencionalnih shema reprezentacije, kao što su pravila ekspertnog sustava ili graf semantičke mreže, u neuronske mreže te obrnuto. Ovakav prijenos naravno mora održavati bitna svojstva izvorne sheme. Na primjer, jedan način transformacije može uključivati konverziju skupa činjenica i pravila iz baze znanja ekspertnog sustava u neuronske mreže. Ovime se efektivno uspostavlja domena prethodnog znanja u mreži što u konačnici može pomoći u olakšavanju procesa učenja, a dodatnu prednost predstavlja brže vrijeme odaziva sustava kao i veća robusnost zbog njegove sposobnosti generalizacije. Suprotan proces predstavljala bi transformacija neuronske mreže u skup pravila i činjenica koje se može iskoristiti za generiranje eksplicitnije, simboličko orijentirane reprezentacije. Ovakav postupak poželjan je u nekoliko situacija. Na primjer, ako se rješava problem koji sadrži veliku količinu informacija neuronske mreže mogu poslužiti kao prvi korak za filtriranje i generalizaciju podataka, a eksplicitnija reprezentacija za dokumentaciju i verifikaciju znanja. On je također koristan kao alat za analizu mreža, pružajući objašnjenje njihovog sadržaja prijevodom u formalna pravila koja su podložnija ljudskoj inspekciji. Iako su danas dostupne brojne metode za izradu ovakvih sustava, još uvijek ne postoji u potpunosti automatizirana transformacija koja se odnosi na opće probleme.⁷⁰
3. Slabo povezani sustavi su, jednostavno rečeno, oni koji razmjenjuju manje dijelove znanja između različitih komponenti, za razliku od prethodnog pristupa koji prebacuje cjelokupne sustave prikaza iz jedne sheme u drugu. Ovo je moguće ostvariti na nekoliko načina kao što je komunikacija putem datoteka, korištenje različitih komponenti za prethodnu i naknadnu obradu informacija itd. Konkretni primjeri bili bi sekvencijalni pristup koji koristi izlazne podatke jednog modula kao ulazne podatke drugog ili pak koprociranje, gdje je nekoliko komponenti istovremeno aktivno te razmjenjuju informacije po potrebi. Slabo povezane sustave je relativno lako implementirati zato što se oslanjaju na postojeće komponente, no zato je potrebno uspostaviti konkretan protokol za razmjenu informacija i za sinkronizaciju aktivnosti individualnih komponenti.⁷¹
4. Čvrsto povezani sustavi uspostavljaju komunikaciju između modula pomoću zajedničke memorije. U ovoj shemi, strukture podataka jedne komponente izravno su dostupne drugoj čime je omogućena njihova brza interakcija. Ovi sustavi često nude

⁷⁰ Ibid.

⁷¹ Ibid.

veliku fleksibilnost dizajna i integracije komponenti, kao i postizanje učinkovitije komunikacije što u konačnici rezultira s većim performansama od slabo povezanih sustava. S druge strane, zbog navedenih karakteristika oni obično posjeduju puno veću kompleksnost s višom razinom međuovisnosti između komponenti. Samim time, ovakve sustave je puno teže razviti.⁷²

5. Potpuno integrirani sustavi, za razliku od prethodnih tipova, koriste zajednički prikaz znanja za sve svoje komponente. Ovakvi sustavi posjeduju ugrađenu sposobnost simboličke i podsimboličke interpretacije i reprezentacije. Iz perspektive prikaza znanja, oni ne moraju posjedovati odvojene komponente za pohranu, manipulaciju i pronalaženje znanja, a razlika između simboličko i podsimboličko orijentiranih pristupa u ovakvim sustavima postaje nejasna. U načelu, potpuno integrirani sustavi posjeduju veće sposobnosti u usporedbi s drugim pristupima. No isto tako imaju puno manji stupanj redundancije što može uzrokovati manju toleranciju kvarova. Također, zbog nedostatka metoda i alata za njihov dizajn i implementaciju, karakterizira ih iznimno visoki stupanj kompleksnosti.

Za kraj bitno je napomenuti kako je razvoj hibridnih arhitektura relativno noviji trend. Implementacija potpuno različitih UI paradigmi unutar jednog sustava još uvijek nije usavršena. Određeni koncepti, poput prethodno opisanih potpuno integriranih sustava, trenutačno nisu dovoljno praktični kako bi ih se koristilo u svakodnevnim situacijama te su za sada ograničeni na istraživačka i eksperimentalna okruženja. Od relevantnijih pristupa trenutačno su najpopularnije arhitekture koje kombiniraju ekspertne sustave s komponentama neuronskih mreža te transformacijska metoda ekstrakcije pravila kao osnova za razmjenu znanja između komponenti.⁷³

4.3.1. ACT-R

ACT-R (*Adaptive Components of Thought-Rational*) je hibridna kognitivna arhitektura i teorijski okvir za simulaciju i razumijevanje ljudskih kognitivnih sposobnosti. Osnovni dijelovi ACT-R sustava su skup perceptivno-motoričkih modula, memorijskih modula, međuspremnik i mehanizam za prepoznavanje uzoraka. Perceptivno-motorički moduli služe kao sučelje između sustava i okoline. Memorijski moduli koriste se za deklarativnu i proceduralnu memoriju, a u oba slučaja uspostavljeni su kao simboličko-konekcioniističke

⁷² Ibid.

⁷³ Ibid.

strukture. Simbolička razina sastoji se od produkcijskih pravila za proceduralnu memoriju ili od komada (*chunks*) za deklarativnu memoriju, a podsimbolička razina izgrađena je u obliku masivno paralelne konekcionističke strukture. Međuspremnici pak služe za privremenu pohranu komunikacije između modula dok se mehanizam za prepoznavanje uzoraka koristi za pronalaženje produkcije u proceduralnoj memoriji koja odgovara trenutnom stanju u međuspremnicima. ACT-R koristi pristup učenja odozgo. Ulazni podaci u obliku perceptualne informacije, cilja ili aktivacije deklarativne memorije pretvaraju se u komade koji se pohranjuju u međuspremnicima, a produkcijski sustav pronalazi pravilo koje najbolje odgovara tom stanju. Podsimbolički parametri zatim se prilagođavaju korištenjem Bayesove vjerojatnosti kako bi se simbolički konstrukti koji su korisni učinili izraženijima. Na taj način komadi koji se često koriste postaju aktivniji te ih se može dohvatiti brže i pouzdanije. Ova arhitektura je primijenjena u velikom broju psiholoških studija i inteligentnih tutorskih sustava, ali konkretna implementacija u području rješavanja problema i zaključivanja još nije uspješno provedena.⁷⁴

4.3.2. CLARION

CLARION (*Connectionist Learning Adaptive Rule Induction ON-line*) je još jedan dobar primjer hibridne arhitekture koja funkcioniра na temelju različitih eksplicitnih i implicitnih procesa te njihove međusobne komunikacije. Njegova arhitektura sadrži četiri memorijska modula ili podsustava, od kojih svaki koristi dualnu simboličko-podsimboličku reprezentaciju: akcijski orijentiran, ne-akcijski orijentiran, motivacijski te metakognitivni podsustav. Prvi je zadužen za reguliranje agentovih akcija dok drugi održava opće znanje sustava (eksplicitno ili implicitno). Treći modul osigurava motivaciju ili poticaj za percepciju, akciju i spoznaju, a metakognitivni podsustav nadzire, usmjerava i mijenja rad ostala tri modula. Svaki od tih modula koristi lokalno-distribuiranu reprezentaciju na način da lokalni dio sadrži eksplicitno znanje, a distribuirani implicitno. CLARION također koristi različite metode učenja. Na primjer, usvajanje implicitnog znanja postiže se raznim metodama učenja s podrškom kao što je *Q*-učenje ili metodama s nadzorom kao što je standardni algoritam s povratnim postupkom, a ono se zatim koristi kako bi se oblikovalo eksplicitnije znanje na višim razinama sustava. Ovo je standardan pristup učenja odozdo, iako je sustav sposoban i za obrnuti postupak. Do danas je ova arhitektura uspješno primijenjena u simulaciji psiholoških

⁷⁴ Duch, W.; Oentaryo, R. J.; Pasquier, M. Cognitive Architectures: Where do we go from here?, u: *Frontiers in Artificial Intelligence and Applications*, vol. 171, uredili Wang, P; Goertzel, B. i Franklin, S. IOS Press, ožujak 2008. Str. 122-136. URL: <http://www.fizyka.umk.pl/publications/kmk/08-AGI.pdf> (20.8.2016.)

podataka, ali i u postupcima složenog sekvencijalnog odlučivanja kao što je navigacija minskim poljem.⁷⁵

4.3.3. LIDA

LIDA (*Learning Intelligent Distribution Agent*) je hibridna arhitektura koja se koristi simboličkom i konekcionističkom organizacijom memorije u različitim modulima za percepciju, radnu memoriju, semantičku memoriju, izbor akcije itd. Sustav izvršava većinu operacija pomoću malih dijelova koda ili programa (*codelets*) koji obavljaju visoko specijalizirane jednostavne zadatke. Veliki naglasak unutar arhitekture stavljen je na perceptualno učenje novih objekata, kategorija, odnosa itd. Ovakvo učenje ima dva osnovna mehanizma, a to su jačanje ili slabljenje aktivacije čvorova kao i stvaranje novih čvorova i veza u perceptivnoj memoriji. Bitnu ulogu također ima i epizodno učenje specifičnih događaja, kao i proceduralno učenje novih akcija i akcijskih sekvenci kojima se izvršavaju novi zadatci. Osnovni princip kojim se sustav koristi je učenje s podrškom. Iako ova arhitektura uspješno naglašava brojne aspekte uma, još uvijek nije uspješno demonstrirala razumijevanje prirodnog jezika ili zdravorazumsko zaključivanje na temelju percepcije.⁷⁶

⁷⁵ Ibid.

⁷⁶ Ibid.

5. ARGUMENTI PROTIV UOI SUSTAVA

Iako je područje umjetne opće inteligencije još uvijek usmjereno prema izradi istinski inteligentnih sustava i agenata, tijekom godina nemali broj znanstvenika i filozofa izrazio je svoje sumnje u ostvarivost takve ideje. Neke od tih kritika bile su usmjerene na relativno trivijalne aspekte, poput toga da su tehnološka ograničenja još uvijek prevelika, ali određeni broj njih cilja na same temelje ideje umjetne inteligencije. Kao i u slučaju prethodnih aspekata ovoga rada, količina relevantnih kritika prevelika je kako bi se sve kvalitetno prikazale, ali moguće je ukratko objasniti nekoliko njih koje su do danas ostale utjecajne.

5.1. UI kao aksiomatski sustav

Ovaj poznati argument najčešće se veže uz teoreme Kurta Gödela iz 1931. godine. Dovoljno je reći kako prema ovim teoremima postoje određene tvrdnje koje formalni sustavi ne mogu dokazati dok ljudi instinktivno vide istinu u njima. Naime, prema ovome viđenju računalni programi su zatvoreni sustavi čija je osnovna zadaća donositi zaključke na temelju prethodno utvrđenih aksioma, a glavni kriterij tih zaključaka je da održavaju istinitost aksioma. Veliki zagovornici ove teze su filozof John Lucas kao i matematičar Roger Penrose. Štoviše, Penrose je izrazito uvjeren kako se ljudsko razumijevanje i uvid ne mogu reducirati na pojam algoritma te se samim time ne mogu oponašati računalima. Iz očitih i prethodno spomenutih razloga, opće inteligentni agent teško može funkcionirati kao aksiomatski sustav. No prema Lucasu, jedina alternativa je izgradnja iracionalnog sustava koji ne slijedi nikakva pravila. Među glavnim kritičarima ovakvog pristupa istaknuli su se Peter Norvig i Stuart Russell koji su tvrdili kako se ovi argumenti odnose samo na ono što je teoretski dokazivo, uz uvjet da postoji beskonačna količina memorije i vremena dok u praksi stvarni strojevi i ljudi imaju ograničene resurse te će imati poteškoća u dokazivanju mnogih teorema. Suština ovakvog pristupa je da nije potrebno dokazati sve kako bi bio inteligentan.⁷⁷

Iako je u određenim krugovima Gödelov argument ostao utjecajan, većina UI znanstvenika smatra kako on ne ograničava mogućnost izgradnje opće inteligentnog sustava. Ono u čemu se slažu je da takav hipotetski sustav ne može, ali i ne mora biti zatvoren. Štoviše, postoji određeni konsenzus kako će istinski umjetno inteligentni agent biti otvoren u sljedećem smislu:

⁷⁷ Philosophy of artificial intelligence. Wikipedia.

URL: http://en.wikipedia.org/wiki/Philosophy_of_artificial_intelligence#Lucas.2C_Penrose_and_G.C3.B6del (20.8.2016.)

- Sustav dobiva znanje s vremena na vrijeme iz svoje okoline, što znači da ne postoji stalni skup aksioma iz kojih se izvode svi zaključci
- Pravila zaključivanja na temelju dostupnog znanja često rezultiraju s uvjerljivim zaključcima, iako oni mogu biti u konfliktu s novim informacijama
- Sustav može posjedovati eksplicitne ili implicitne kontradikcije između dijelova znanja ili zaključaka⁷⁸

5.2. UI i Turingov stroj

Ova kritika relativno je slična prethodnoj te se odnosi na tezu kako je UI sustav ekvivalent Turingovom stroju. Jednostavno rečeno, Turingov stroj je uređaj koji sadrži konačan skup stanja od kojih je samo jedno početno te jedno ili više konačnih stanja. On također sadrži i vrpcu s nizom ulaznih simbola. U svakom koraku procesa računanja trenutno stanje unutar stroja i simbol koji se nalazi pod glavom za čitanje određuju hoće li biti promjene stanja, izlaznog simbola na vrpici ili kretanja glave po traci. Kompletan proces počinje na nekom početnom stanju te završava u jednom od završnih stanja. U ovome smislu, Turingov stroj odgovara konceptu algoritma. Sažetak ovog argumenta pružio je Lucas koji kaže: „Naša ideja stroja je upravo to da njegovo ponašanje potpuno određuje način na koji je napravljen te ulazni podražaji. Tu ne postoji mogućnost njegovog vlastitog djelovanja. Uz određeni oblik izgradnje i određeni unos podataka, on mora djelovati na neki specifičan način.“⁷⁹

Ovakav argument je varijanta onoga što je sam Turing nazivao *Lady Lovelace's Objection* koji u suštini kaže da kako bi računalo riješilo problem, ono mora slijediti predodređeni algoritam. No u svakodnevnom životu problem obično nije dobro definirani skup već konkretna instanca koja se tretira individualno. U ljudskom umu obično nema unaprijed određenog algoritma za obradu svih takvih problema određenog tipa na jedinstven način. Umjesto toga, oni se obrađuju od slučaja do slučaja s bilo kojom metodom koje se um prisjeti ili formira u tom trenutku. U konačnici, zbog toga što se čini kako ljudski um ne slijedi ovakav unaprijed određeni proces, ljudi poput Lucasa ovo vide kao temeljno ograničenje umjetne inteligencije. Do danas je predloženo nekoliko rješenja ovog problema,

⁷⁸ Wang, P. Three Fundamental Misconceptions of Artificial Intelligence, u: *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 19, br.3, Taylor & Francis Inc., rujan 2007. Str. 249-268.

URL: http://cis-linux1.temple.edu/~wangp/Publication/AI_Misconceptions.pdf (20.8.2016.)

⁷⁹ Lucas, J. R. Minds, machines and Gödel, u: *Philosophy*, vol. 36, 1961. Str. 112–127.

URL: https://www.openstarts.units.it/dspace/bitstream/10077/5466/1/Lucas1_E%26P_V_2003_1.pdf (20.8.2016.)

no u svojoj osnovi ova kritika još uvijek postoji. Na primjer, Wang naglašava kako UI agent ne mora nužno posjedovati algoritam za svaku moguću situaciju već da njegovo ponašanje usmjerava određena vrsta meta-algoritma koji mu omogućuje učenje novih rješenja kao i usavršavanje starijih. Prema njemu ovo nije ista stvar zato što algoritam, ili Turingov stroj, mora biti striktno definiran u odnosu na određeni proces rješavanja problema, uz konkretno određene ulazno-izlazne parove. No Penrose smatra kako ovdje ne postoji nikakva distinkcija te kako takav inteligentni agent podliježe istoj kritici.⁸⁰

5.3. UI i pitanje semantike

Vjerojatno najpoznatija kritika umjetne inteligencije do danas odnosi se na pitanje istinskog razumijevanja pojmova te se uglavnom veže uz Johna Searlea i njegov argument Kineske sobe koji je u načelu jednostavan. On uključuje čovjeka koji se nalazi sam u prostoriji s knjigom i komadom papira. Kroz jedan otvor u prostoriji on prima papir s pitanjima u obliku kineskih znakova, zatim koristeći upute u knjizi on ispisuje odgovore također u obliku kineskih znakova te ih prosljeđuje kroz drugi otvor u prostoriji. Poanta ovog misaonog eksperimenta je da ljudima izvan prostorije izgleda kao da čovjek u njoj priča kineski iako on ne posjeduje nikakvo znanje iz tog područja. Konkretno, Searle smatra kako manipuliranje simbolima bez njihovog razumijevanja ne podrazumijeva um ili inteligenciju. Jednostavnije rečeno, sintaksa nije dovoljna za semantiku što znači da računalni programi koji su striktno definirani svojom formalnom sintaktičkom strukturom ne mogu istinski razumjeti simbole kojima manipuliraju.

Česta zablude je kako se ova kritika isključivo odnosi na koncept hipoteze sustava fizičkih simbola, odnosno na tradicionalnu simboličku UI. Doduše, između nje i konekcionista pristupa postoji jedna fundamentalna razlika. Naime, kod simboličkog pristupa osnovne sintaktičke jedinice su računalni tokeni, recimo LISP atomi, kojima se manipulira. Kod neronskih mreža, ove osnovne jedinice su individualni neuroni i veze. Razlika je u tome što su LISP atomi ujedno i osnovne semantičke ili reprezentacijske jedinice, odnosno njihova osnovna funkcija je označavanje nečega. S druge strane, osnovni semantički mehanizam kod mreža je distribuirani uzorak aktivnosti na razini svih neurona i njihovih veza. Konkretno, u ovome slučaju sintaktička razina nalazi se ispod reprezentacijske razine. Uostalom, iz ove razlike proizlazi sam naziv podsimboličkog područja. No u konačnici,

⁸⁰ Wang, P. Three Fundamental Misconceptions of Artificial Intelligence, u: *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 19, br.3, Taylor & Francis Inc., rujan 2007. Str. 249-268.
URL: http://cis-linux1.temple.edu/~wangp/Publication/AI_Misconceptions.pdf (20.8.2016.)

Searleov argument općenito je usmjeren protiv računalnih sustava, a konekcionistički modeli implementirani su upravo kao računalni programi.⁸¹

Tijekom godina razvijeno je nekoliko odgovora na ovu kritiku. Jedan od glavnih je princip uzemljenja simbola (*symbol grounding*). Konkretno, radi se o načelu povezivanja računalnih simbola s njihovim stvarnim instancama u fizičkom svijetu. Ovakav pristup naglašava kako jedini način posjedovanja istinskih reprezentacija može doći kroz senzo-motoričku interakciju s okolinom. Na primjer, prava računalna reprezentacija pojma psa može se jedino ostvariti u prisustvu pravog fizičkog psa u okolini. Na ovaj način ona posjeduje svoju eksternu referencu te samim time i značenje.⁸² Iako je ovaj pristup u određenoj mjeri prihvaćen, za Searlea on ne predstavlja nikakav napredak. Prema njemu mentalna stanja i svijest proizlaze iz stvarnih fizičkih i kemijskih svojstava ljudskog mozga, odnosno iz posebnih uzročnih svojstava mozga i neurona iz kojih se rađa um.⁸³

5.4. Ostale kritike

Zadnji dio ovog poglavlja ne odnosi se toliko na kritike filozofske naravi već na općenite probleme unutar tri prethodno opisana pristupa izrade umjetnih opće inteligentnih sustava. Za simbolički pristup često se naglašava kako sadrži brojne vrijedne ideje te kako je do danas postigao mnoge zanimljive rezultate. No isto tako, ovaj pristup još uvijek se nije pokazao sposobnim za izradu struktura potrebnih kako bi se dobila opća inteligencija nalik ljudskoj. Činjenica je kako je ljudska manipulacija simbola nastala evolucijski na temelju jednostavnijih procesa percepcije i motivacije. Brojni znanstvenici smatraju kako raskidanje ove veze nema smisla te kako nikada neće doprinjeti izradi opće inteligentnih agenata već u najboljem slučaju samo korisnim alatima za rješavanje problema.

S druge strane, za konekcionističke ideje često se navodi kako su pogrešno fokusirane na podlogu ljudske inteligencije. Pristaše ovakvih kritika smatraju kako je jedino bitan kognitivni softver uma, a ne niskorazinski hardver na kojemu se on koristi. Oni naglašavaju kako mozak posjeduje kompleksnu arhitekturu koju je evolucija specifično razvila kako bi podržala napredno simboličko zaključivanje i druge aspekte ljudske opće inteligencije. Ono

⁸¹ Chalmers, D. J. Subsymbolic Computation and the Chinese Room, u: *The Symbolic and Connectionist Paradigms: Closing the Gap*, uredio Dinsmore, J. Erlbaum, 1992. Str. 25-48.

URL: <http://consc.net/papers/subsymbolic.pdf> (20.8.2016.)

⁸² Ibid.

⁸³ Philosophy of artificial intelligence. Wikipedia.

URL: http://en.wikipedia.org/wiki/Philosophy_of_artificial_intelligence#Searle.27s_Chinese_room (20.8.2016.)

što je bitno za stvaranje takve inteligencije je posjedovanje prave arhitekture za obradu informacija, a ne mehanizam kojim se ona implementira.

I na kraju, kao generalni argument protiv hibridnih arhitektura uglavnom se navodi kako povezivanje nekoliko neadekvatnih sustava ne će odjednom proizvesti adekvatan sustav. Ovdje se naglašava da mozak koristi jedinstvenu infrastrukturu iz dobrog razloga jer kada pokušate povezati kvalitativno različite komponente dobijete krhki sustav koji se ne može dobro prilagoditi zato što različite komponente ne mogu raditi zajedno s punom fleksibilnošću. Kao najveće zamjerke hibridnih sustava ističu se njihova neelegantnost te kršenje najjednostavnijeg principa Ockhamove oštrice.⁸⁴

Konačna istina je kako većina ovih i drugih kritika nemaju neki preveliki utjecaj na istraživanja unutar područja umjetne inteligencije. Ovaj stav najbolje su opisali Russell i Norvig koji kažu kako većina UI istraživača i znanstvenika ne mari za hipotezu jake umjetne inteligencije te dok god program radi njih ne zanima da li se to zove simulacija inteligencije ili stvarna inteligencija.⁸⁵

⁸⁴ Goertzel, B. Artificial General Intelligence: Concept, State of the Art, and Future Prospects, u: *Journal of Artificial General Intelligence*, vol. 5, br. 1, 15. ožujka 2014. Str. 1-48.

URL:<https://www.degruyter.com/downloadpdf/j/jagi.2014.5.issue-1/jagi-2014-0001/jagi-2014-0001.xml> (20.8.2016.)

⁸⁵ Russell, S. J.; Norvig, P. Artificial Intelligence: A Modern Approach. Prentice-Hall, Englewood Cliffs, 1995.

URL:<http://stpk.cs.rtu.lv/sites/all/files/stpk/materiali/MI/Artificial%20Intelligence%20A%20Modern%20Approach.pdf> (20.8.2016.)

6. Zaključak

Bez obzira na dosadašnje neuspjehe u ostvarenju ideje opće inteligentnih sustava, prevladavajuće mišljenje unutar UI zajednice je kako se radi o pitanju vremena. Mogućnost inteligentnih programa koji su sposobni generalizirati svoje znanje na širokom području smatra se inženjerskim problemom. Ako se prihvati takva pretpostavka, onda se izrada UOI sustava doista svodi na izradu dovoljno efikasnog algoritma ili meta-algoritma koji će usmjeravati procese učenja, komunikacije i interakcije s okolinom. U ovakvom kontekstu gdje se inteligencija u velikoj mjeri izjednačava sa sposobnošću razumijevanja i rješavanja različitih problema, vrlo je vjerojatno kako će sadašnji pristupi daljnjim postupnim napretcima u konačnici rezultirati s pravim UOI sustavima. Hoće li to biti kroz nekoliko destljeća ili stoljeća, teško je predvidjeti. Naravno, uvijek postoji mogućnost da su ljudi poput Johna Searlea u pravu te kako ljudska inteligencija uvelike ovisi o raznim svojstvima koja još uvijek dovoljno ne razumijemo, a možda nikada i nećemo. Ako su takva predviđanja točna, moguće je kako će ideal umjetne inteligencije zauvijek ostati neostvariv.

7. Literatura

Arel, I.; Rose, D.; Coop, R. DeSTIN: A Scalable Deep Learning Architecture with Application to High-Dimensional Robust Pattern Recognition. AAAI Fall Symposium Series, siječanj 2009. URL: <http://web.eecs.utk.edu/~itamar/Papers/BICA2009.pdf> (20.8.2016.)

Artificial intelligence, Evolutionary computing. Encyclopaedia Britannica.
URL: <http://www.britannica.com/EBchecked/topic/37146/artificial-intelligence-AI/219093/Evolutionary-computing> (20.8.2016.)

Bayesian probability. Wikipedia.
URL: https://en.wikipedia.org/wiki/Bayesian_probability (20.8.2016.)

Bounded rationality. Wikipedia.
URL: https://en.wikipedia.org/wiki/Bounded_rationality (20.8.2016.)

Chalmers, D. J. Subsymbolic Computation and the Chinese Room, u: *The Symbolic and Connectionist Paradigms: Closing the Gap*, uredio Dinsmore, J. Erlbaum, 1992. Str. 25-48.
URL: <http://consc.net/papers/subsymbolic.pdf> (20.8.2016.)

Control theory. Wikipedia. URL: https://en.wikipedia.org/wiki/Control_theory (20.8.2016.)

Duch, W.; Oentaryo, R. J.; Pasquier, M. Cognitive Architectures: Where do we go from here?, u: *Frontiers in Artificial Intelligence and Applications*, vol. 171, uredili Wang, P; Goertzel, B. i Franklin, S. IOS Press, ožujak 2008. Str. 122-136. URL: <http://www.fizyka.umk.pl/publications/kmk/08-AGI.pdf> (20.8.2016.)

Expert system. Wikipedia. URL: http://en.wikipedia.org/wiki/Expert_system (20.8.2016.)

Explanation-based learning. Wikipedia. URL: https://en.wikipedia.org/wiki/Explanation-based_learning (20.8.2016.)

Genetic algorithm. Wikipedia.
URL: https://en.wikipedia.org/wiki/Genetic_algorithm (20.8.2016.)

Goertzel, B. Artificial General Intelligence: Concept, State of the Art, and Future Prospects, u: *Journal of Artificial General Intelligence*, vol. 5, br. 1, 15. ožujka 2014. Str. 1-48.
URL: <https://www.degruyter.com/downloadpdf/j/jagi.2014.5.issue-1/jagi-2014-0001/jagi-2014-0001.xml> (20.8.2016.)

Hebbian theory. Wikipedia. URL: https://en.wikipedia.org/wiki/Hebbian_theory (20.8.2016.)

History of artificial intelligence. Wikipedia.
URL: http://en.wikipedia.org/wiki/History_of_artificial_intelligence#Dartmouth_Conference_1956:_the_birt_h_of_AI (20.8.2016.)

Intelligent agent. Wikipedia. URL: http://en.wikipedia.org/wiki/Intelligent_agent (20.8.2016.)

Kieras, D. E.; Meyer, D. E. An Overview of the EPIC Architecture for Cognition and Performance With Application to Human-Computer Interaction, u: *Human-Computer Interaction*, vol. 12, br. 4, prosinac 1997. Str. 391-438.
URL: <http://umich.edu/~bcalab/documents/KierasMeyer1997.pdf> (20.8.2016.)

Kurfess, F. J. Integrating symbol-oriented and subsymbolic reasoning methods into hybrid systems, u: *From Synapses to Rules - Discovering Symbolic Rules from Neural Processed Data*, uredili Apolloni, B. i Kurfess, F. J. Kluwer Academic Publishers, 2002. Str. 275–292.
URL: http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1016&context=csse_fac (20.8.2016.)

Lucas, J. R. Minds, machines and Gödel, u: *Philosophy*, vol. 36, 1961. Str. 112–127.
URL: https://www.openstarts.units.it/dspace/bitstream/10077/5466/1/Lucas1_E%26P_V_2003_1.pdf (20.8.2016.)

Mitchell, T. M.; Thrun, S. B. Learning Analytically and Inductively, u: *Mind Matters: A Tribute to Alan Newell*, uredili Steier, D. M. i Mitchell, T. M. Erlbaum, 1996. Str. 85-111.
URL: https://www.ri.cmu.edu/pub_files/pub1/mitchell_tom_1996_1/mitchell_tom_1996_1.pdf (20.8.2016.)

Moravec's paradox. Wikipedia.
URL: http://en.wikipedia.org/wiki/Moravec's_paradox (20.8.2016.)

Newell, A.; Simon, H. A. Computer Science as Empirical Inquiry: Symbols and Search, u: *Communications of the ACM*, vol. 19, br. 3, ožujak 1976. Str. 113-126.
URL: <http://www.cs.utexas.edu/~kuipers/readings/Newell+Simon-cacm-76.pdf> (20.8.2016.)

Nilsson, N. J. Human-Level Artificial Intelligence? Be Serious!, u: *AI Magazine*, vol. 26, br. 4, 2005. Str. 68-75.
URL: <http://www.aaai.org/ojs/index.php/aimagazine/article/view/1850/1748> (20.8.2016.)

Pennachin, C.; Goertzel, B. Contemporary Approaches to Artificial General Intelligence. AGIRI - Artificial General Intelligence Research Institute, 1405 Bernerd Place, Rockville, MD 20851, SAD, 2007.
URL: http://www.springer.com/cda/content/document/cda_downloadaddocument/9783540237334-c1.pdf?SGWID=0-0-45-330089-p43950079 (20.8.2016.)

Philosophy of artificial intelligence. Wikipedia.
URL: http://en.wikipedia.org/wiki/Philosophy_of_artificial_intelligence#Lucas.2C_Penrose_and_G.C3.B6del (20.8.2016.)

Production system (computer science). Wikipedia.
URL: [https://en.wikipedia.org/wiki/Production_system_\(computer_science\)](https://en.wikipedia.org/wiki/Production_system_(computer_science)) (20.8.2016.)

Rodney A. Brooks. Elephants Don't Play Chess, u: *Robotics and Autonomous Systems*, vol. 6, br. 1-2, lipanj 1990. Str. 3-15. URL: <http://people.csail.mit.edu/brooks/papers/elephants.pdf> (20.8.2016.)

Russell, S. J.; Norvig, P. Artificial Intelligence: A Modern Approach. Prentice-Hall, Englewood Cliffs, 1995.
URL: <http://stpk.cs.rtu.lv/sites/all/files/stpk/materiali/MI/Artificial%20Intelligence%20A%20Modern%20Approach.pdf> (20.8.2016.)

Semantic network. Wikipedia.
URL: https://en.wikipedia.org/wiki/Semantic_network (20.8.2016.)

Sun, R. Artificial Intelligence: Connectionist and Symbolic Approaches, u: *International Encyclopedia of the Social and Behavioral Sciences*, uredili Smelser, N. J. I Baltes, P. B. Pergamon/Elsevier, Oxford, 2001. Str. 783-789.

URL: <http://www.cogsci.rpi.edu/~rsun/sun.encyc01.pdf> (20.8.2016.)

Sun, R. Supplementing Neural Reinforcement Learning with Symbolic Methods, u: *Hybrid Neural Systems*. Springer-Verlag, London, UK, prosinac 1998. Str. 333-347. URL: www.cogsci.rpi.edu/~rsun/sun.hybrid99.ps (20.8.2016.)

Wang, P. Artificial General Intelligence and Classical Neural Network, u: *Proceedings of the IEEE International Conference of Granular Computing*. IEEE Computational Intelligence Society, Atlanta, Georgia, svibanj 2006. Str. 130-135.

URL: <http://cis-linux1.temple.edu/~pwang/Publication/AGI-CNN.pdf> (20.8.2016.)

Wang, P. Embodiment: Does a laptop have a body? Department of Computer and Information Sciences, Temple University, Philadelphia, SAD, 2009.

URL: http://agi-conference.org/2009/papers/paper_5.pdf (20.8.2016.)

Wang, P. Non-Axiomatic Logic: A Model of Intelligent Reasoning. Department of Computer and Information Sciences, Temple University Philadelphia, PA 19122, SAD, 2013.

URL: <http://cis-linux1.temple.edu/~pwang/Publication/Diagnosis.pdf> (20.8.2016.)

Wang, P. The Logic of Intelligence. Department of Computer and Information Sciences, Temple University Philadelphia, PA 19122, SAD, 2006.

URL: http://www.cogsci.indiana.edu/farg/peiwang/PUBLICATION/wang.logic_intelligence.pdf (20.8.2016.)

Wang, P. Three Fundamental Misconceptions of Artificial Intelligence, u: *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 19, br.3, Taylor & Francis Inc., rujan 2007. Str. 249-268.

URL: http://cis-linux1.temple.edu/~wangp/Publication/AI_Misconceptions.pdf (20.8.2016.)

Wang, P. What Do You Mean by "AI"? u: *Proceedings of the 2008 conference on Artificial General Intelligence*, Memphis, Tennessee, ožujak 2008., str. 362-373.

URL: http://cis-linux1.temple.edu/~wangp/Publication/AI_Definitions.pdf (20.8.2016.)